

PT Gene encoding variable region of anti-human influenza A type virus  
 PT antibody - useful for prodn. of artificial antibodies  
 PS Example 4: Page 36-38; 42pp; English.  
 CC This sequence represents the scFv-PP fusion polypeptide. This sequence  
 CC is an anti-human influenza A type virus antibody. The antibody  
 CC recognises the stem region of the haemagglutinin (HA) molecule of the  
 CC H1N1 and H2N2 subtypes of human influenza A type virus, and shows  
 CC neutralisation activity against these two subtypes. The antibody shows  
 CC no recognition of the H3N2 subtype. Artificial antibodies (such as this)  
 CC and polypeptides are useful in the diagnosis and treatment of human  
 CC influenza. As the antibodies recognise the stem region of the HA  
 CC molecule, the influenza virus will be recognised even if the HA molecule  
 CC changes. This provides an advantage over current vaccines, as the virus  
 CC periodically alters its HA molecule.  
 SQ Sequence 392 AA;

Query Match 29.6%; Score 72; DB 16; Length 392;  
 Best Local Similarity 32.5%; Pred. No. 3.57e+01;  
 Matches 13; Conservative 14; Mismatches 11; Indels 2; Gaps 2;

Db 139 svlssggsgsgsgsgsgsgsgs-di-elsgsgsgsgs176  
 1 GIGISEADGCGKGMARCKSIATGDIQAALGSGSIATGDI 40

RESULT 9  
 ID R13498 standard; Protein: 459 AA.

AC R13498;  
 DT 25-OCT-1991 (first entry)  
 DE P.dentriticans COB G.  
 KW cob gene; corrinoid; descobalocorrinoid; cor igene.  
 OS Pseudomonas dentritificans.  
 PN W09111518-A.  
 PD 08-AUG-1991.  
 PF 30-JAN-1991; F00054.  
 PR 31-JAN-1990; FR-001137.  
 PA (RHON) RHONE-POULENC BIOCH.  
 PI Blanche F, Meron B, Crouzet J, Debussche L, Lévy-Schill S;  
 PI Thibaut D;  
 PI WPI; 91-252650/34.  
 DR N-PSDB; Q13285.  
 PT New polypeptide(s) involved in cobalamin and cobamide  
 PT biosynthesis - and DNA encoding them, for amplification of  
 PT cobalamin, esp. coenzyme B12 prodn.  
 PS Claim 17; Fig 16; 29pp; French.  
 CC This sequence corresponds to one of 24 polypeptides obtained from  
 CC P.dentriticans and implicated in the biosynthesis of cobalamines  
 CC and/or cobamides. COBG is specifically involved in the conversion  
 CC of precorrin-3 to 5'-deoxy-5'-adenosyl cobyrinic acid a,c-diamide.  
 CC It is encoded by part of the 8.7kb EcoRI-EcoRI fragment of plasmid  
 CC pXJ367. The plasmid was isolated from a P.dentriticans genomic DNA  
 CC bank constructed in vector pXJ59.  
 CC See Q13284-Q13288.  
 SQ Sequence 459 AA;

Query Match 29.6%; Score 72; DB 3; Length 459;  
 Best Local Similarity 35.0%; Pred. No. 3.57e+01;  
 Matches 14; Conservative 9; Mismatches 16; Indels 1; Gaps 1;

Db 67 gnglielta-rgnqlrqlsaasvprlagaigdaetaiae 105  
 1 GIGISEADGCGKGMARCKSIATGDIQAALGSGSIATGDI 40

RESULT 10  
 ID W09221 standard; Protein: 936 AA.  
 AC W09221;  
 DT 30-MAR-1997 (first entry)  
 DE SELPDK-CSI polymer.  
 KW SELPDK-CSI; protein polymer; silk; fibroin; elastin;  
 KW repeat sequence; cleavage site; collagenase; crosslinking;  
 KW adhesive; tissue sealant; biocompatible film;  
 OS Synthetic.

FH Key Location/Qualifiers  
 FT peptide 34..39  
 FT region /note= "Silk fibroin peptide (W09206, claim 4)"  
 FT region /note= "Silk fibroin peptide repeat sequence"  
 FT peptide 46..50  
 FT cleavage\_site 51..56  
 FT region /note= "Double collagenase cleavage site (W09215)"  
 FT region 57..91  
 FT misc\_difference 93  
 FT /note= "Residue with functional group for crosslinking"  
 FT region 97..111  
 FT region /note= "Elastin peptide repeat sequence"  
 FT region 34..111  
 FT /note= "Block monomer sequence"  
 FT region 34..903  
 FT /note= "Block of 15 identical monomer sequences"  
 FT region 904..915  
 FT /note= "Silk fibroin peptide repeat sequence"

PS Example 3, Page 34; 103pp; English.  
 CC This SELPDK-CSI polymer sequence (mol.wt. 76,389) has been  
 CC constructed from monomers composed of repeat units derived from  
 CC silk fibroin (W09206) and elastin (W09207) (claim 4). The monomer  
 CC (W09220) has been amplified by multicopy cloning in Escherichia  
 CC coli (using plasmid pM0370) to produce a recombinant protein polymer  
 CC backbone with lys functional groups for subsequent polyfunctional  
 CC chemical crosslinking. A double collagenase cleavage site (W09215)  
 CC has been inserted to promote in vivo resorption. The crosslinked  
 CC polymer product forms a strongly adherent tissue adhesive or sealant.  
 CC The polymer may be used to seal defects in vessel walls, e.g. artery,  
 CC vein, capillary, lung, dura or colon, to increase tissue mass, or  
 CC to produce biocompatible films for in vivo use. The sealant has the  
 CC biocompatibility of fibrin glues, but sets more quickly, and has  
 CC greater strength. Polymers of this type are readily prepared by  
 CC recombinant methods, are easy to administer, and are gradually  
 CC resorbed.  
 SQ Sequence 936 AA;

Query Match 29.6%; Score 72; DB 20; Length 936;  
 Best Local Similarity 21.1%; Pred. No. 3.57e+01;  
 Matches 8; Conservative 19; Mismatches 10; Indels 1; Gaps 1;

Db 87 gvyvpgagsgagagsgv-gvplgplgpygyvgyv 123  
 1 GIGISEADGCGKGMARCKSIATGDIQAALGSGSIATGDI 38

RESULT 11  
 ID W06600 standard; Protein: 344 AA.  
 AC W06600;  
 DT 30-MAR-1997 (first entry)  
 DE Hypersensitive response elicitor protein.  
 KW Hypersensitive response; elicitor; Pseudomonas solanacearum; plant;  
 KW disease-resistance; Escherichia coli; infiltration; virus;  
 KW bacterium; fungus; pathogen; biological control agent.  
 OS Pseudomonas solanacearum.  
 PN W09639602-A1.  
 PD 19-DEC-1996.  
 PF 05-JUN-1996; 008819.



```

KW Outer membrane protein-106; OMP106; vaccine; immune response;
KV Cytotoxic antibody; Moraxella catarrhalis; primer; probe.
OS Moraxella catarrhalis.
PN W09741731-Al.
PF 13-NOV-1997.
PR 28-APR-1997; U07679.
PA (ANTE-) ANTEK BIOLOGICS INC.
PI Pliosin L Tucker K.
PW WP1: 97-558601/51.
DR N-PSSD: T86522.
PT Outer membrane protein, OMP106, of Moraxella catarrhalis - used in vaccines for producing immune responses against M. catarrhalis
PS Disclosure: Page 58: 78pp; English.
CC This is a partial sequence of an amino terminal peptide fragment of a novel outer membrane protein-106 (OMP106) polypeptide. The OMP106 is an outer membrane polypeptide of Moraxella catarrhalis, a haemagglutinating cultivar. The encoding DNA can be used as a 5' primer for PCR amplification of a full length OMP106 DNA. The DNA fragment can also be used as a probe for screening M. catarrhalis genomic libraries for OMP106 polypeptide coding sequences. The encoded peptide fragment can specifically bind to an antibody which mediates complement killing of M. catarrhalis. The OMP106 polypeptide, and its peptide fragments can be used in vaccines and antigenic compositions. They can also be used for producing an immune response in an animal against M. catarrhalis.
SQ Sequence 24 AA:

Query Match          60.1%; Score 146; DB 28; Length 24;
Best Local Similarity 100.0%; Pred.No. 2,61e+05;
Matches   24; Conservative    0; Mismatches     0; Indels      0; Gaps       0;

Db         1 eadggkgsanargdkslaiqdlg 24
           ||||| | | | | | | | | | | |
Qy        6 EADGGKGANARGDKSLAIQDLIAQ 29

RESULT            3
ID               W04505 standard; Protein; 1992 AA.
AC              W04505;
DT             25-JAN-1997 (first entry)
DE           Morxella 200 kba outer membrane protein.
KW           Outer membrane protein; OMP; Immunogen; vaccine; otitis media; diagnosis.
OS           Mycobacterium catarrhalis strain 4223.
PN             WO9634960-A1.
PD             07-NOV-1996.
PE             29-APR-1996; CA0264.
PR             01-MAY-1995; US-431718.
PR             07-JUN-1995; US-478370.
PR             26-MAR-1996; US-621944.
PP           ((CONN)) CONNAUGHT LAB LTD.
PQ           Chong P., Harkness RE, Klein MH, Loosmore SM, Sasaki K;
PW             WIJ: 96-506162/50.
DR             N-PSSDB: T38740.
PT           Moraxella outer membrane protein - useful as immunogen in protective vaccine and for diagnosis
PS           Claim 14; Fig 6; 109pp; English.
CC           An approx. 200 kDa outer membrane protein (W04505) can be isolated from Moraxella catarrhalis otitis media strain 4223 by electroprecipitation, or expressed from a gene (see also T38740) obt'd. from a strain 4223 genomic library. Natural or recombinant outer membrane protein is useful as an immunogen to protect against infection by Moraxella, esp. M. catarrhalis. It can also be used to detect antibodies, esp. for differential diagnosis between bacteria that cause similar symptoms, and also useful as a carrier for other antigens and used to raise antitumour antibodies for conjugation to therapeutic agents.
SQ           Sequence 1992 AA:

Query Match          44.0%; Score 107; DB 20; Length 1992;
Best Local Similarity 50.0%; Pred. No. 5.55e+02;
Matches   20; Conservative    11; Mismatches     7; Indels      2; Gaps       2;

```

```

Db      20   iaieqngprirsgakadgdgdraiaiganaaggsgataig 59
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      2   IGISEADGGR-GG-ANARGDKSIAIGDIAALGSQSIAIG 39

RESULT         4
ID      RS6490 standard; Protein; 278 AA.
AC      RS6490;
DT      23-MAR-1995 (first entry)
DE      TATA-binding protein-associated factor dTAFlit40.
KW      TATA-binding protein associated factor; dTAFlit40; screening;
KM      diagnostic; therapeutic; gene transcription regulation.
OS      Drosophila.
PN      W09417087-A.
PD      04-AUG-1994.
PF      28-JAN-1994; US-01114.
PR      28-JAN-1993; US-013412.
PR      30-JUN-1993; US-087119.
PA      (REGC ) UNIV CALIFORNIA.
PI      Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI      Tjian R, Wang E, Weinzierl ROJ;
PI      WPI: 94-264019/32.
DR      P-PDB: C70728.
PT      TATA-binding protein associated protein factors - and
PT      corresponding nucleotide sequence and deriv. antibodies, useful
PS      in screening diagnostics and therapeutics
PS      Disclosure; Page 92-93; 180pp; English.
CC      The TATA-binding protein associated factor dTAFlit40 (including
CC      specific antibodies and fusion products) are used in drug screening,
CC      diagnostics and therapeutics. They are used in the development of
CC      specific biochemical assays for screening compounds that agonise or
CC      antagonise selected transcription factors involved in regulating
SQ      gene expression associated with human pathology.
SQ      Sequence     278 AA:

Query Match              30.9%; Score 75; DB 11; Length 278;
Best Local Similarity 31.6%; Pred. No. 2.10e+01;
Matches    12; Conservative    12; Mismatches    14; Indels      0; Gaps      0.

Db      198   vggsgggggggqewkseeatggagqltknewdsaaayg 235
          :|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      2   IGISEADGGKGGAMARGDKSIAGDIAALGSQSIAIG 39

RESULT         5
ID      W25025 standard; Protein; 278 AA.
AC      W25025;
DT      08-OCT-1997 (first entry)
DE      TATA-binding protein associated factor; dTAFlit40.
KW      TATA-binding protein associated factor; TAF; nuclear protein;
KM      RNA polymerase transcription; TATA-binding protein; TBP;
KM      initiation.
OS      Drosophila sp.
PN      US6537686-A.
PD      10-JUN-1997.
PF      28-JAN-1993; 013412.
PR      28-JAN-1993; US-013412.
PR      30-JUN-1993; US-087119.
PR      28-JUN-1994; US-188582.
PR      09-MAY-1996; US-646715.
PA      (REGC ) UNIV CALIFORNIA.
PI      Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;
PI      Tjian R, Wang E, Weinzierl ROJ;
PI      WPI: 97-319113/29.
DR      N-P-SDB: T79601.
PT      Nucleic acids encoding human TATA-binding protein associated factor
PT      (TAF) peptide(s) - for production of recombinant peptide(s), used
PT      for modulating transcription of TAFs
PS      Example 1; Column 63-64; 86pp; English.
CC      W25025 represents TATA-binding protein associated factor (TAF)
CC      polypeptide. dTAFlit40 (mol. weight 40KD). TAF peptides derived
CC      from dTAFlit10 alpha, dTAFlit30 beta, dTAFlit40, dTAFlit60, dTAFlit80,
```







Best Local Similarity 57.1%; Pred. No. 5.05e+01;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 65 GAILGGR 71

QY 1 GTVLGGR 7

RESULT 15  
ID 018604 PRELIMINARY; PRT; 287 AA.  
AC 018604;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL PROTEIN C4B9.3.  
LN C4B9.3  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MCMURRAY A.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A., CRAXTON M.,  
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HARKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R., SMAILDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WIKTINSON-SPROUT J., WOHLDMAN P.;  
RL NATURE 368:32-38(1994).  
DR EMBL; 273424; E348268; -.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 287 AA; 32138 MW; 259B3473 CRC32;

Query Match 81.3%; Score 39; DB 5; Length 287;

Best Local Similarity 71.4%; Pred. No. 8.47e+01;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 217 STVLGGR 223

QY 1 GTVLGGR 7

Search completed: Wed Oct 28 11:32:03 1998  
Job time : 38 secs.

RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U97013, G193831; -  
KW AMINOCYL-TRNA SYNTHASE.  
SQ SEQUENCE 968 AA; 106781 MW; 30BC629D CRC32;

Query Match 85.4%; Score 41; DB 5; Length 968;  
Best Local Similarity 100.0%; Pred. No. 2.99e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 929 TVLGK 934  
|||  
QY 2 TVLGK 7

RESULT 11  
ID 055766 PRELIMINARY; PRT; 1026 AA.  
AC 055766;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DE 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE PUTATIVE DNA-DEPENDENT RNA POLYMERASE LARGEST SUBUNIT.  
DE CHILLO IRIDESCANT VIRUS (CIV) (INSECT IRIDESCANT VIRUS TYPE 6).  
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; IRIDOVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA BARR U., TIDONA C.A., DARAI G.;  
RL VIRUS GENES 0:0-0(1997).  
DR EMBL: AF003534; G2738450; -  
SQ SEQUENCE 1026 AA; 116963 MW; AB524CDB CRC32;

Query Match 85.4%; Score 41; DB 14; Length 1026;  
Best Local Similarity 85.7%; Pred. No. 2.99e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 695 GTVLNGK 701  
|||  
QY 1 GTVLGK 7

RESULT 12  
ID 089506 PRELIMINARY; PRT; 1051 AA.  
AC 089506;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE DNA-DEPENDENT RNA POLYMERASE LARGEST SUBUNIT HOMOLOG.  
DE CHILLO IRIDESCANT VIRUS (CIV) (INSECT IRIDESCANT VIRUS TYPE 6).  
OS AND IRIDESCANT VIRUS TYPE 6.  
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; IRIDOVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 95213160.  
RA SONNTAG K.C., SCHNITZLER P., JANSSEN W., DARAI G.;  
RL INTERVIROLOGY 37:287-297(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 93118242.  
RA SONNTAG K.C., DARAI G.;  
RL VIRUS GENES 6:333-342(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 94353641.  
RA SONNTAG K.C., SCHNITZLER P., KOONIN E.V., DARAI G.;  
RL VIRUS GENES 8:151-158(1994).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 94229206.  
RA SCHNITZLER P., SONNTAG K.C., MULLER M., JANSSEN W., BUGERT J.J.,  
RA KOONIN E.V., DARAI G.;  
RL J. GEN. VIROL. 75:1557-1567(1994).

DR EMBL: S75674; G861468; -  
DR EMBL: M81388; G553222; -  
KW PFAM: PF00623; RNA\_POL\_A.  
DR DNA-DIRECTED RNA POLYMERASE.  
SQ SEQUENCE 1051 AA; 120348 MW; 332C7F46 CRC32;

Query Match 85.4%; Score 41; DB 14; Length 1051;  
Best Local Similarity 85.7%; Pred. No. 2.99e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 694 GTVLNGK 700  
|||  
QY 1 GTVLGK 7

RESULT 13  
ID 099129 PRELIMINARY; PRT; 1150 AA.  
AC 099129;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
DE SMUT FUNGUS (CLONE PML-1) MYPI (CLONE PML-1).  
GN MYPI.  
OS USTILAGO MAYDIS (SMUT FUNGUS).  
OC EUKARYOTA; FUNGI; BASIDIOMYCOTINA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-518;  
RL MEDLINE; 96109597.  
RA GIASSON L., KRONSTAD J.W.;  
RL GENETICS 141:491-501(1995).  
DR EMBL: L33919; G886446; -  
SQ SEQUENCE 1150 AA; 123387 MW; 68647042 CRC32;

Query Match 85.4%; Score 41; DB 3; Length 1150;  
Best Local Similarity 100.0%; Pred. No. 2.99e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1019 GTVLGK 1024  
|||  
QY 1 GTVLGK 6

RESULT 14  
ID 051337 PRELIMINARY; PRT; 328 AA.  
AC 051337;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE PROTIPOPROTEIN DIACYLGLYCERYL TRANSFERASE (LGT).  
GN BB0362.  
OS BORRELLIA BURGENSIS (LYME DISEASE SPIROCHETE).  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; SPIROCHETES; SPIROCHAETALES;  
OC SPIROCHAETACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 35210 / B31;  
RX MEDLINE; 96065943.  
RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,  
RA LATHIGRA R., WHITE O., KETCHUM K.A., DOBSON R., HICKER E.K., GINN M.,  
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D., HANSON M.,  
RA PETERSON J., KERLAVAGE A.R., OUCKENBUSH J., SALZBERG S., BOWMAN J.,  
RA VAN VUOT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,  
RA UTTERBACK T., WATTHEY L., McDONALD L., ARTIACH P., ROBERTS K., HATCH B.,  
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,  
RA SMITH H.O., VENTER J.C.;  
RL NATURE 390:580-586(1997).  
DR EMBL: AE001142; G2688274; -  
DR TIGR: BB0362; -  
KW TRANSFERASE; LIPOPROTEIN  
SQ SEQUENCE 328 AA; 37595 MW; 7EEDD5 CRC32;

Query Match 83.3%; Score 40; DB 2; Length 328;

RC STRAIN-BRISTOL N2;  
RA BECK C., FRONICK W., WILSON R.;  
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AF047663; G2911883; -  
SQ SEQUENCE 178 AA; 16110 MW; 5C9B440D CRC32;

Query Match 85.4%; Score 41; DB 5; Length 178;  
Best Local Similarity 100.0%; Pred. No. 2,99e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 153 GTVLGG 158  
|||||  
OY 1 GTVLGG 6

RESULT 7  
ID 059496 PRELIMINARY; PRT; 202 AA.  
AC 059496;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE 202AA LONG HYPOTHETICAL PROTEIN.  
GN PHCY051  
OS PYROCOCUS HORIKOSHII.  
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-OT3;  
RA KAMARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOTAYAMA A., NAGAI Y.,  
RA SAKAI M., OGURA K., OTSUKA R., NAKAZAWA H., TAKAMURA M., OHFUKU Y.,  
RA FUNAHASHII T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
RA AOKI K., YOSHIZAWA T., NAKAMURA Y., MASUCHI Y., SHIZUO H., KIKUCHI H.;  
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AB009526; D1028936; -  
SQ SEQUENCE 202 AA; 21299 MW; 8711A332 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 202;  
Best Local Similarity 100.0%; Pred. No. 2,99e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 141 GTVLGG 146  
|||||  
OY 1 GTVLGG 6

RESULT 8  
ID 045082 PRELIMINARY; PRT; 449 AA.  
AC 045082;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE MOBP.  
GN MOBP.  
OS BURKHOLDERIA CEPACIA (PSEUDOMONAS CEPACIA).  
OC PLASMID PMOP.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC PSEUDOMONADACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PC701;  
RX MEDLINE: 96425865.  
RA SAINT C.P., ROMAS P.;  
RL MICROBIOLOGY 143:2407-2416(1996).  
DR EMBL: U29532; G903977; -  
DR PFAM: PF00083; sugar\_tr.  
KW PLASMID.  
SQ SEQUENCE 449 AA; 48305 MW; CE2C7FEF CRC32;

Query Match 85.4%; Score 41; DB 2; Length 449;  
Best Local Similarity 100.0%; Pred. No. 2,99e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 395 GTVLGG 400  
|||||  
OY 1 GTVLGG 6

RESULT 9  
ID 060266 PRELIMINARY; PRT; 933 AA.  
AC 060266;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE KIAA0511 PROTEIN (FRAGMENT).  
GN KIAA0511.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN.  
RA NAGASE T., ISHIMAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,  
RA OHARA O.;  
RL DNA RES. 5:31-39(1998).  
DR EMBL: AB011083; D1026367; -  
DR PROSITE: PS00452; GUANYLATE\_CYCLASES; 2.  
KW LYSASE.  
FT NON\_TER  
SQ SEQUENCE 933 AA; 105485 MW; E7D6D757 CRC32;

Query Match 85.4%; Score 41; DB 4; Length 933;  
Best Local Similarity 100.0%; Pred. No. 2,99e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 207 GTVLGG 212  
|||||  
OY 1 GTVLGG 6

RESULT 10  
ID 001541 PRELIMINARY; PRT; 968 AA.  
AC 001541;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE STIMULAR TO AMINOACYL-TRNA SYNTHETASE.  
GN P28H1.3.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTING J., LLOYD C.,  
RA MCNUTRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RL NATURE 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA DU Z., LE T.T.;  
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE DNA FOR PLCA/PRA OPERON.  
 OC LISTERIA SELEGIERI.  
 OC PROKARYOTA, FIRMICUTES; REGULAR ASPOROGENOUS ROD; UNCERTAIN.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-SLCC379;  
 RA LAMPIDIS R., KREFT J.;  
 RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL: X97014; E242618; -;  
 DR PRAM: PF00596; Aldolase II.  
 SQ SEQUENCE 233 AA; 25981 MW; 7FAAF950 CRC32;  
 Query Match 89.6%; Score 43; DB 2; Length 233;  
 Best Local Similarity 85.7%; Pred. No. 1.01e+01;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 69 GSVLGGK 75  
 QY 1 GTVLGGK 7

RESULT 3  
 ID 053394 PRELIMINARY; PRT; 484 AA.  
 AC 053394;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DE 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
 DE PGSS-FAMILY PROTEIN.  
 GN MYO16.44C.  
 OS MYCOBACTERIUM TUBERCULOSIS.  
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA OLIVER K., HARRIS D.;  
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;  
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE: 96181548.  
 RA PHILIPPE W.J., POULET S., EIGLMEIER K., PASCOPELLA L.;  
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
 RA COLE S.T.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).  
 DR EMBL: AL021841; E1251176; -;  
 SQ SEQUENCE 484 AA; 37491 MW; 21D7718E CRC32;  
 Query Match 89.6%; Score 43; DB 2; Length 484;  
 Best Local Similarity 85.7%; Pred. No. 1.01e+01;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 145 GSVLGGK 151  
 QY 1 GTVLGGK 7

RESULT 4  
 ID 035408 PRELIMINARY; PRT; 858 AA.  
 AC 035408;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE PROMININ.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE-KIDNEY;  
 RX MEDLINE: 98024147.  
 RA WEIGMANN A., CORBEIL D., HELLMIG A., HUTNER W.B.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 94:12425-12430(1997).  
 DR EMBL: AF026269; G2559004; -;  
 SQ SEQUENCE 858 AA; 96222 MW; 89F0244E CRC32;  
 Query Match 87.5%; Score 42; DB 11; Length 858;  
 Best Local Similarity 71.4%; Pred. No. 1.75e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 226 GSVLGGK 232  
 QY 1 GTVLGGK 7

RESULT 5  
 ID 054990 PRELIMINARY; PRT; 867 AA.  
 AC 054990;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE AC133 ANTIGEN HOMOLOG.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE: 98052559.  
 RA MIRAGLIA S., GODFREY W., YIN A.H., ATKINS K., WANKKE R., HOLDEN J.T.,  
 RA BRAY R.A., WALLER E.K., BUCK D.W.;  
 RL BLOOD 90:5013-5021(1997).  
 DR EMBL: AF039663; G2789658; -;  
 SQ SEQUENCE 867 AA; 97112 MW; EC2C10AD CRC32;  
 Query Match 87.5%; Score 42; DB 11; Length 867;  
 Best Local Similarity 71.4%; Pred. No. 1.75e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 235 GSVLGGK 241  
 QY 1 GTVLGGK 7

RESULT 6  
 ID 045197 PRELIMINARY; PRT; 178 AA.  
 AC 045197;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
 DE W09G12.6 PROTEIN.  
 GN W09G12.6.  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTIA; RHABDITIDA.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERRS M., BONFIELD J.,  
 RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
 RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
 RA HAWKINS T., HILLIER L., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
 RA KIRSTEN J., LAISTER N., LAURENCE P., PARSONS J., PERCY C.,  
 RA MCMORRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
 RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,  
 RA SONNHAMMER E., STADEN R., SULISTON J., THIERRY-MIEG J., THOMAS K.,  
 RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
 RA WIKINSON-SPROAT J., WOHLDMAN P.;  
 RL NATURE 368:32-38(1994).  
 RN [2]  
 RC SEQUENCE FROM N.A.

\*\*\*\*\*  
M P E R E I A  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Oct 28 11:31:25 1998; Maspar time 4.29 Seconds

Tabular output not generated. 81.235 Million cell updates/sec

Title: >US-08-968-685-11

Description: (1-7) from US08968685.pep

Sequence: 1 GTVLGK 7

Scoring table: PAM 150

Gap 15

Searched: 165420 seqs, 49795644 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

sptrembl6

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rndent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 20.564; Variance 21.000; scale 0.979

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	43	89.6	230	2	025735	HYPOTHETICAL 25.7 KD P	1.01e+01
2	43	89.6	233	2	007687	DNA FOR PLCA/PRA OPER	1.01e+01
3	43	89.6	484	2	053394	PGRS-FAMILY PROTEIN.	1.01e+01
4	42	87.5	856	11	035408	PROMITIN	1.75e+01
5	42	87.5	867	11	054990	AC133 ANTIGEN HOMOLOG.	1.75e+01
6	41	85.4	178	5	045197	W09G12.6 PROTEIN.	1.75e+01
7	41	85.4	202	1	059496	202AA LONG HYPOTHETICA	2.99e+01
8	41	85.4	449	2	045082	MOBP.	2.99e+01
9	41	85.4	933	4	060266	KIAA0511 PROTEIN (FRAG	2.99e+01
10	41	85.4	968	5	001541	STIMILAR TO AMINOACIL-T	2.99e+01
11	41	85.4	1028	14	055766	PUTATIVE DNA-DEPENDENT	2.99e+01
12	41	85.4	1051	14	089506	DNA-DEPENDENT RNA POLYM	2.99e+01
13	41	85.4	1150	3	099129	SMUT FUNGUS (CLONE PM1	2.99e+01
14	40	83.3	328	2	051337	PROLIPROTEIN DIACYLG	5.05e+01
15	39	81.3	287	5	018604	HYPOTHETICAL PROTEIN C	8.47e+01
16	39	81.3	298	6	028547	PROGESTERONE RECEPTOR	8.47e+01
17	39	81.3	301	2	007398	MAV301.	8.47e+01
18	39	81.3	337	2	054817	DAUNORUBICIN-DOXORUBIC	8.47e+01
19	39	81.3	354	2	044328	MCCA.	8.47e+01
20	39	81.3	360	13	042374	PROGESTERONE RECEPTOR	8.47e+01

ID	Query Match	Score	DB	Length	Best Local Similarity	Pred. No.	Score	DB	Length	Best Local Similarity	Pred. No.
21	39	81.3	443	14	098518	GENOME, PARTIAL SEQUEN	8.47e+01				
22	39	81.3	519	14	074230	CORE PROTEIN.	8.47e+01				
23	39	81.3	570	10	040406	PHYTOENE DESATURASE.	8.47e+01				
24	39	81.3	571	10	041849	PHYTOENE DESATURASE.	8.47e+01				
25	39	81.3	582	5	061587	ACETYLCHOLINESTERASE B	8.47e+01				
26	39	81.3	923	11	063449	PROGESTERONE RECEPTOR	8.47e+01				
27	39	81.3	1109	6	000756	GLYCOCEN-ASSOCIATED SE	8.47e+01				
28	39	81.3	1128	1	052009	REP1 PROTEIN.	8.47e+01				
29	39	81.3	1205	2	031208	17 KDA ANTIGEN (FRAGME	1.40e+02				
30	38	79.2	131	2	052637	17 KDA ANTIGEN (FRAGME	1.40e+02				
31	38	79.2	137	2	031065	17 KDA ANTIGEN (FRAGME	1.40e+02				
32	38	79.2	154	2	053154	(CLONE PRB F15E 1), 5'	1.40e+02				
33	38	79.2	267	2	046982	2-OXO-HEPT-3-ENE-1,7-D	1.40e+02				
34	38	79.2	372	2	045622	SAC OPERON RELATED REG	1.40e+02				
35	38	79.2	407	2	053879	HYPOTHETICAL 40.0 KD P	1.40e+02				
36	38	79.2	414	11	063662	RAT 3-METHYLCOLANTHRE	1.40e+02				
37	38	79.2	422	2	P74896	QAH/OAS SULFHYDRYLASE.	1.40e+02				
38	38	79.2	439	2	034921	YTOT.	1.40e+02				
39	38	79.2	445	2	006171	HYPOTHETICAL 46.1 KD P	1.40e+02				
40	38	79.2	456	5	062376	T106.6.	1.40e+02				
41	38	79.2	530	11	P97866	UDP-GLUCURONOSYLTRANSF	1.40e+02				
42	38	79.2	631	2	051879	PUTATIVE CHROMOSOME RE	1.40e+02				
43	38	79.2	646	14	084460	GENOME, PARTIAL SEQUEN	1.40e+02				
44	38	79.2	1320	14	006359	150 KD PROTEIN.	1.40e+02				
45	38	79.2	2153	5	094246	CODED FOR BY C. ELEGAN	1.40e+02				

#### ALIGNMENTS

RESULT 1  
ID 025735. PRELIMINARY: PRT: 230 AA.

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 25.7 KD PROTEIN.

GN HPI107  
OS HELICOBACTER PYLORI (CAMPILOBACTER PYLORI).  
OC PROKARYOTA: GRACILICUTES; SCOTOBACTERIA.  
OC AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-26695:

RX MEDLINE: 97394467.

RA TOME J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,

RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,

RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,

RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GRODEK A.,

RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKER E.K.,

RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,

RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATNEY L., WALLIN E.,

RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,

RA VENTER J.C.,

RL NATURE 388:539-547(1997).

DR EMBL: AE000617; G2314267; -.

DR TIGR: HPI107; -.

KW HYPOTHETICAL PROTEIN.

SQ SEQUENCE 230 AA; 25697 MW; FB44C6CD CRC32;

Query Match 89.6%; Score 43; DB 2; Length 230;  
Best Local Similarity 85.7%; Pred. No. 1.01e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 82 GTVLGK 88  
Oy 1 GTVLGK 7

RESULT 2  
ID 007687. PRELIMINARY: PRT: 233 AA.  
AC 007687;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)



-----



DR PIR: JX0152; JX0152.  
 DR PROSITE; PS00342; MICROBODIES\_CTER; 1.  
 DR PROSITE; PS00677; DAO; 1.  
 KW OXIDOREDUCTASE; FLAVOPROTEIN; FAD.  
 FT MOD\_RES ?1 ?1 BLOCKED.  
 FT NP\_BIND 5 19 FAD (ADP PART) (POTENTIAL).  
 FT ACT\_SITE 242 242 BY SIMILARITY.  
 FT ACT\_SITE 327 327 BY SIMILARITY.  
 FT SITE 359 361 MICROBODI TARGETING SIGNAL (POTENTIAL).  
 SQ SEQUENCE 361 AA: 39696 MW: 4BD27642 CRC32;  
  
 Query Match 81.3%; Score 39; DB 1; Length 361;  
 Best Local Similarity 83.3%; Pred. No. 3.65e+01;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 251 GTVLGG 256  
 11:111  
 QY 1 GTVLGG 6

Search completed: Wed Oct 28 11:31:06 1998  
 Job time : 9 secs.

DR PIR; S40005; S40005.  
 DR HSP; P20231; 1A0.  
 DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 KW HYDROLASE; SERINE PROTEASE; SIGNAL; ZYMOMEN; DIGESTION;  
 KW MULTIGENE FAMILY.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT PROPER 19 48 ACTIVATION PEPTIDE.  
 FT CHAIN 49 275 TRYPsin 4.  
 FT ACT\_SITE 89 89 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 134 134 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 230 230 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 74 90 BY SIMILARITY.  
 FT DISULFID 199 215 BY SIMILARITY.  
 FT DISULFID 226 250 BY SIMILARITY.  
 FT SITE 224 224 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
 SQ SEQUENCE 275 AA; 29695 MW; 8196C784 CRC32;  
 Query Match 81.3%; Score 39; DB 1; Length 275;  
 Best Local Similarity 71.4%; Pred. No. 3.65e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Db 76 GSVLGGK 82  
 1 GTVLGGK 7  
 RESULT 13 STANDARD; PRT; 299 AA.  
 ID ISPA.ECOLI  
 AC P22939;  
 DT 01-AUG-1991 (REL. 19, CREATED)  
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE GERANYLTRANSTRANSFERASE (EC 2.5.1.10) (FARNESYL-DIPHOSPHATE SYNTHASE)  
 DE (FPP SYNTHASE).  
 GN ISPA.  
 OS ESCHERICHIA COLI.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE; 91210228.  
 RA FUJISAKI S., HARA H., NISHIMURA Y., HORIUCHI K., NISHINO T.;  
 RL J. BIOCHEM. 108:995-1000(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA ROBERTS D., ALLEN E., ARAUJO R., APARICIO A., CHUNG E., DAVIS K.,  
 DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KORDI O.,  
 LEW H., LIN D., NAMATH A., OEFNER P., SCHRAMM S., DAVIS R.W.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- CATALYTIC ACTIVITY: GERANYL DIPHOSPHATE + ISOPRENYL DIPHOSPHATE  
 = PYROPHOSPHATE + TRANS,TRANS-FARNESYL DIPHOSPHATE.  
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER FPP  
 SYNTHETASE, GGPP SYNTHETASE AND HEXAPRENYL PYROPHOSPHATE  
 SYNTHETASE.  
 CC EMBL; D00694; G216584; -;  
 DR EMBL; AE000146; G1786623; -;  
 DR EMBL; U82664; G1773105; -;  
 DR PIR; J00665; J00665.  
 DR SWISS-2DPAGE; P22939; COLI.  
 DR ECOGENE; E310508; ISPA.  
 DR PROSITE; PS00444; POLYPRENYL\_SYNTHET\_2; 1.  
 DR PROSITE; PS00723; POLYPRENYL\_SYNTHET\_1; 1.  
 KW TRANSFERASE; ISOPRENE BIOSYNTHESIS.  
 SQ SEQUENCE 299 AA; 32159 MW; 2EB4409F CRC32;  
 Query Match 81.3%; Score 39; DB 1; Length 299;

Best Local Similarity 71.4%; Pred. No. 3.65e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Db 39 GALLGGK 45  
 1 GTVLGGK 7  
 RESULT 14 STANDARD; PRT; 321 AA.  
 ID MKRD\_KLEPN  
 AC P21648;  
 DT 01-MAY-1991 (REL. 18, CREATED)  
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE FIMBRIA ADHESIN PROTEIN PRECURSOR.  
 GN MKRD.  
 OS KLEBSIELLA PNEUMONIAE.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-1A565;  
 RX MEDLINE; 91100388.  
 RA ALLEN B.L., GERLACH G.-F., CLEGG S.;  
 RL J. BACTERIOL. 173:916-920(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89155420.  
 RA GERLACH G.-F., CLEGG S., ALLEN B.L.;  
 RL J. BACTERIOL. 171:1262-1270(1989).  
 CC -1- SUBCELLULAR LOCATION: FIMBRIA.  
 CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.  
 DR EMBL; M55912; G149238; -;  
 DR EMBL; M24556; G511858; -;  
 DR PIR; E39142; E39142.  
 DR PIR; B32801; B32801.  
 KW SIGNAL; FIMBRIA.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 321 FIMBRIA ADHESIN PROTEIN.  
 FT CONFLICT 113 113 Y->V (IN REF. 2).  
 SQ SEQUENCE 321 AA; 33969 MW; CFC2F728 CRC32;  
 Query Match 81.3%; Score 39; DB 1; Length 321;  
 Best Local Similarity 83.3%; Pred. No. 3.65e+01;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 179 TILGGK 184  
 2 TVLGGK 7  
 RESULT 15 STANDARD; PRT; 361 AA.  
 ID OXDA\_FUSSO  
 AC P24552;  
 DT 01-MAR-1992 (REL. 21, CREATED)  
 DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE D-AMINO ACID OXIDASE (EC 1.4.3.3) (DAMOXY) (DAO) (DAAO).  
 OS FUSARIUM SOLANI (SUBSP. PIS).  
 CC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-M-0718 / FERM P-2688.  
 RX MEDLINE; 91210207.  
 RA ISOGAI T., ONO H., ISHITANI Y., KOJO H., UEDA Y., KOHSAKA M.;  
 RL J. BIOCHEM. 108:1063-1069(1990).  
 CC -1- FUNCTION: THIS ENZYME CAN EFFECTIVELY CONVERT CEPHALOSPORIN C  
 INTO 7-BETA-(5-CARBOXY-5-OXOPENTANAMIDO)-CEPHALOSPORINIC ACID.  
 CC -1- CATALYTIC ACTIVITY: A D-AMINO ACID + H(2)O + O(2) = A 2-OXO-ACID +  
 NH(3) + H(2)O(2)  
 CC -1- COFACTOR: FAD FLAVOPROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE DAMOX/DASOX FAMILY.  
 DR EMBL; D00809; G217923; -;

RP SEQUENCE OF 1-129 FROM N.A.  
 RC STRAIN-168;  
 RX MEDLINE: 94321318.  
 RA MIREL D.B., LAUER P., CHAMBERLIN M.J.;  
 RL J. BACTERIOL. 176:4492-4500(1994).  
 CC -1- SIMILARITY: BELONGS TO THE FLAGELLA BASAL BODY ROD PROTEINS  
 CC FAMILY.  
 CC EMBL: U56901; G1762338; -  
 DR EMBL: L14437; G451871; -  
 DR SUBTILIST; BG10401; FLAG.  
 DR PROSITE; PS00588; FLAGELLA\_BB\_ROD; 1.  
 KW FLAGELLA.  
 SQ SEQUENCE 507 AA; 54355 MW; 5C9952C8 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 507;  
 Best Local Similarity 85.7%; Pred. No. 1.23e+01;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 256 GTVLGK 262  
 111111  
 1 GTVLGK 7

RESULT 10  
 ID CYA3\_RAT STANDARD; PRT; 1144 AA.  
 AC P21932;  
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE ADENYLATE CYCLASE, OLFACTIVE TYPE (EC 4.6.1.1) (TYPE III) (ATP  
 DE PYROPHOSPHATE-LYASE) (ADENYLYL CYCLASE).  
 GN ADCY3.  
 OS RATUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-OLFACTORY EPITHELIUM;  
 RX MEDLINE: 91075227.  
 RA BAKAIXAR H.A., REED R.R.;  
 RL SCIENCE 250:1403-1406(1990).  
 CC -1- FUNCTION: MEDIATE ODORANT DETECTION (POSSIBLY) VIA MODULATION  
 CC OF INTRACELLULAR CAMP CONCENTRATION.  
 CC -1- CATALYTIC ACTIVITY: ATP - 3',5'-CYCLIC AMP + PYROPHOSPHATE.  
 CC -1- ENZYME REGULATION: ACTIVATED BY CA(2+)/CALMODULIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: CILIA OF THE OLFACTORY EPITHELIUM.  
 CC -1- INDUCTION: BY STIMULATORY G PROTEIN (GS), FORSKOLIN AND ALFA(-).  
 CC -1- SIMILARITY: THE TWO CYTOPLASMIC DOMAINS ARE SIMILAR TO ONE  
 CC ANOTHER AND TO THE CATALYTIC DOMAINS OF GUANYLATE CYCLASES.  
 DR EMBL: M55075; G202715; -  
 DR PIR: A39833; A39833.  
 DR PROSITE; PS00452; GUANYLATE CYCLASES; 2.  
 KW LYASE; CAMP SYNTHESIS; TRANSMEMBRANE; GLYCOPROTEIN; DUPLICATION;  
 KW OLFACTIN.  
 FT DOMAIN 1 79  
 FT TRANSMEM 80 100 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 105 125 POTENTIAL.  
 FT TRANSMEM 139 159 POTENTIAL.  
 FT TRANSMEM 173 193 POTENTIAL.  
 FT TRANSMEM 226 246 POTENTIAL.  
 FT TRANSMEM 381 401 POTENTIAL.  
 FT DOMAIN 402 630 POTENTIAL.  
 FT TRANSMEM 631 651 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 662 682 POTENTIAL.  
 FT TRANSMEM 706 726 POTENTIAL.  
 FT TRANSMEM 755 775 POTENTIAL.  
 FT TRANSMEM 777 797 POTENTIAL.  
 FT TRANSMEM 833 853 POTENTIAL.  
 FT DOMAIN 854 1144 POTENTIAL.  
 FT CARBOHYD 734 734 CTIOPASMIC (POTENTIAL).  
 SQ SEQUENCE 1144 AA; 128935 MW; CF0C3130 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 1144;  
 Best Local Similarity 100.0%; Pred. No. 1.23e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 418 GTVLGK 423  
 111111  
 1 GTVLGK 6

RESULT 11  
 ID LYOA\_HUMAN STANDARD; PRT; 117 AA.  
 AC P04211;  
 DT 20-MAR-1987 (REL. 04, CREATED)  
 DT 20-MAR-1987 (REL. 04, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE IG LAMBDA CHAIN V REGION PRECURSOR (4A).  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 85014122.  
 RA ANDERSON M.L.M., SZAJNERT M.F., KAPLAN J.C., MCCOLL L.,  
 RA YOUNG B.D.;  
 RL NUCLEIC ACIDS RES. 12:6647-6661(1984).  
 DR PIR: A01993; LVH2.  
 DR HSSP; P01703; IMFA.  
 KW IMMUNOGLOBULIN V REGION; SIGNAL.  
 FT SIGNAL 1 20  
 FT CHAIN 21 117 IG LAMBDA CHAIN V REGION (4A).  
 FT DOMAIN 21 42 FRAMEWORK 1, DETERMINING 1.  
 FT DOMAIN 43 55 FRAMEWORK 2, DETERMINING 1.  
 FT DOMAIN 56 71 FRAMEWORK 2, DETERMINING 2.  
 FT DOMAIN 72 78 FRAMEWORK 3, DETERMINING 3.  
 FT DOMAIN 79 110 FRAMEWORK 3, DETERMINING 3.  
 FT DOMAIN 111 117 COMPLEMENTARITY-DETERMINING 3.  
 FT DISULFD 42 110 BY SIMILARITY.  
 FT NON\_TER 117 117  
 FT SEQUENCE 117 AA; 12380 MW; 21B7EF45 CRC32;

Query Match 81.3%; Score 39; DB 1; Length 117;  
 Best Local Similarity 71.4%; Pred. No. 3.65e+01;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 86 GSLLGK 92  
 111111  
 1 GTVLGK 7

RESULT 12  
 ID TRY4\_ANOGA STANDARD; PRT; 275 AA.  
 AC P35038;  
 DT 01-FEB-1994 (REL. 28, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE TRYPSIN 4 PRECURSOR (EC 3.4.21.4).  
 GN TRY4.  
 OS ANOPHELES GAMBIAE (AFRICAN MALARIA MOSQUITO).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SUKOKO; TISSUE-MIDGUT;  
 RX MEDLINE: 9332779.  
 RA MEDLER H.M., CRAMPTON J.M., DELLA TORRE A., SINDEN R., CRISANTI A.,  
 RL EMBO J. 12:2891-2900(1993).  
 CC -1- FUNCTION: MAJOR FUNCTION MAY BE TO AID IN DIGESTION OF THE BLOOD  
 CC MEAL.  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- TISSUE SPECIFICITY: MIDGUT.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DR EMBL: Z22930; G410326; -

OY 1 GTVLGG 6  
 RESULT 7  
 ID CATEB-RAT STANDARD; PRT: 339 AA.  
 AC P00787;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CATHESIN B PRECURSOR (EC 3.4.22.1) (CATHESPIN B1) (RSG-2).  
 GN CTSB.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAFOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-MAMMARY GLAND;  
 RX MEDLINE; 95094788.  
 RA GENETTE R.S., MOOIBROEK M., WONG K., WONG P., TENNISWOOD M.;  
 RL EUR. J. BIOCHEM. 226:311-321(1994).  
 RN [2]  
 RP SEQUENCE OF 69-339 FROM N.A.  
 RX SAN SEGUNDO B., CHAN S.J., STEINER D.F.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 82:2320-2324(1985).  
 RN [3]  
 RP SEQUENCE OF 80-126 AND 129-333.  
 RC TISSUE-LIVER.  
 RX MEDLINE; 83221657.  
 RA TAKIO K., TOMATARI T., KATUNUMA N., TELLER D.C., TITANI K.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 80:3666-3670(1983).  
 RN [4]  
 RP PROCESSING.  
 RX MEDLINE; 92348471.  
 RA ROWAN A.D., MASON P., MACH L., MORT J.S.;  
 RL J. BIOL. CHEM. 267:15993-15999(1992).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RX MEDLINE; 95197558.  
 RA JIA Z., HASNAIN S., HIRAMA T., LEE X., MORT J.S., TO R., HUBER C.P.;  
 RL J. BIOL. CHEM. 270:5527-5533(1995).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 18-339.  
 RX MEDLINE; 96311276.  
 RA CYCLER M., SIVARAMAN J., GROCHULSKI P., COULOMBE R., STORER A.C.,  
 MORT J.S.;  
 RL STRUCTURE 4:405-416(1996).  
 CC -1- FUNCTION: THIOL PROTEASE WHICH IS BELIEVED TO PARTICIPATE IN  
 INTRACELLULAR DEGRADATION AND TURNOVER OF PROTEINS. HAS ALSO  
 BEEN IMPLICATED IN TUMOR INVASION AND METASTASIS.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSES PROTEINS, WITH A SPECIFICITY  
 RESEMBLING THAT OF PAPAIN.  
 CC -1- SUBUNIT: DIMER OF A HEAVY CHAIN AND A LIGHT CHAIN CROSS-LINKED  
 BY A DISULFIDE BOND.  
 CC -1- SUBCELLULAR LOCATION: LYSOSOMAL.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE  
 PAPAIN FAMILY OF THIOL PROTEASES.  
 CC EMBL; X82396; E123172; -;  
 DR EMBL; M1305; G203648; -;  
 DR PIR; A00977; KHRFB.  
 DR PDB; 1CRJ; 07-DEC-95.  
 DR PDB; 1MR; 11-JAN-97.  
 DR PDB; 1CTE; 31-JUL-95.  
 DR PDB; 1THE; 10-MAR-96.  
 DR PROSITE; PS00139; THIOL PROTEASE\_CYS; 1.  
 DR PROSITE; PS00639; THIOL PROTEASE\_HIS; 1.  
 DR PROSITE; PS00640; THIOL PROTEASE\_ASN; 1.  
 DR HYDROLASE; THIOL PROTEASE; LYSOSOME; GLYCOPROTEIN; ZMOGEN; SIGNAL;  
 KM 3D-STRUCTURE.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT PROPEP 18 79 ACTIVATION PEPTIDE.  
 FT CHAIN\* 80 333 CATHESPIN B.  
 FT\*

FT CHAIN 80 126 CATHESPIN B LIGHT CHAIN.  
 FT CHAIN 129 333 CATHESPIN B HEAVY CHAIN.  
 FT PROPEP 334 339  
 FT ACT\_SITE 108 108  
 FT ACT\_SITE 278 278  
 FT ACT\_SITE 298 298  
 FT DISULFID 93 122 BY SIMILARITY.  
 FT DISULFID 105 150 BY SIMILARITY.  
 FT DISULFID 141 207 BY SIMILARITY.  
 FT DISULFID 142 146 BY SIMILARITY.  
 FT DISULFID 179 211 BY SIMILARITY.  
 FT DISULFID 187 198 BY SIMILARITY.  
 FT CARBOHYD 192 192  
 FT VARIANT 302 302  
 FT CONFLICT 159 159  
 FT SEQUENCE 339 AA; 37470 MW; 12DBA727 CRC32;  
 SQ  
 Query Match 85.4%; Score 41; DB 1; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 1.23e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 60 GTVLGG 65  
 OY 1 GTVLGG 6  
 RESULT 8  
 ID Y294\_MYCPN STANDARD; PRT: 475 AA.  
 AC P75366;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL PROTEIN MG294 HOMOLOG.  
 OS MYCOPLASMA PNEUMONIAE.  
 OC PROKARYOTA; TENERICTES; MOLICUTES; MYCOPLASMA; MYCOPLASMALES;  
 OC MYCOPLASMATACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 29342 / M129;  
 RX MEDLINE; 97105885.  
 RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,  
 RA HERMANN R.;  
 RL NUCLEIC ACIDS RES. 24:4420-4449(1996).  
 CC -1- SIMILARITY: SOME, TO E. COLI Y1HN.  
 DR EMBL; AE000041; G1674105; -;  
 KM HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 475 AA; 52731 MW; CB05E707 CRC32;  
 Query Match 85.4%; Score 41; DB 1; Length 475;  
 Best Local Similarity 100.0%; Pred. No. 1.23e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 289 GTVLGG 294  
 OY 1 GTVLGG 6  
 RESULT 9  
 ID FLGK\_BACSU STANDARD; PRT: 507 AA.  
 AC P39810;  
 DT 01-FEB-1995 (REL. 31, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE FLAGELLAR HOOK-ASSOCIATED PROTEIN 1 (HAP1).  
 GN FLGK.  
 OS BACILLUS SUBTILIS.  
 OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168;  
 RA SOLDO B., LAZAREVIC V., MAUEL C., KARAMATA D.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENEBANK/DDSI DATA BANKS.  
 RN [2]

CC -1- SIMILARITY: TO M.JANNASCHII M0678.  
 DR EMBL: U67563; G1591852; -.  
 DR TIGR: M01221; -.  
 KM HYPOTHETICAL PROTEIN; TRANSMEMBRANE.  
 FT TRANSMEM 128 148 POTENTIAL.  
 FT TRANSMEM 177 197 POTENTIAL.  
 FT TRANSMEM 210 230 POTENTIAL.  
 FT TRANSMEM 236 256 POTENTIAL.  
 FT TRANSMEM 261 281 POTENTIAL.  
 SQ SEQUENCE 299 AA; 32538 MW; 8E4A08E7 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 299;  
 Best Local Similarity 85.7%; Pred. No. 1.23e+01;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

161 GTVLGDK 167  
 1 GTVLGGR 7

RESULT 5 STANDARD; PRT; 301 AA.  
 AC P21135;  
 DT 01-MAY-1991 (REL. 18, CREATED)  
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
 DE CUT2 PROTEIN.  
 GN CUT2  
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972;  
 RX MEDLINE; 90367123.  
 RA UZAWA S., SAMEJIMA I., HIRANO T., TANAKA K., YANAGIDA M.;  
 RL EMBL: 621913-925(1990);  
 DR EMBL: M57350; G173375; -.  
 DR PIR: B35694; B35694.  
 KM REPEAT.  
 FT REPEAT 250 260  
 FT REPEAT 270 280  
 SQ SEQUENCE 301 AA; 32854 MW; 40206F69 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 301;  
 Best Local Similarity 100.0%; Pred. No. 1.23e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 53 TVLGGR 58  
 1 GTVLGGR 7  
 2 TVLGGR 7

RESULT 6 STANDARD; PRT; 339 AA.  
 AC P10605;  
 DT 01-JUL-1989 (REL. 11, CREATED)  
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE CATHEPSIN B PRECURSOR (EC 3.4.22.1) (CATHEPSIN B1).  
 GN CTSB.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 91190267.  
 RA QIAN F., FRANKFATER A., CHAN S.-J., STEINER D.F.;  
 RL DNA CELL BIOL. 10:159-168(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 91032179.  
 RA FERRARA M., WOJCICKI F., RHAISSI H., MORDIER S., ROUX M.-P., BECHET D.;  
 FEBS LETT. 273:195-199(1990).

RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87017021.  
 RA CHAN S.-J., SAN SEGUNDO B., MCCORMICK M.B., STEINER D.F.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:7721-7725(1986).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92082172.  
 RA QIAN F., FRANKFATER A., STEINER D.F., BAJROWSKI A.S., CHAN S.-J.;  
 RL ANTICANCER RES. 11:1445-1451(1991).  
 RN [5]  
 RP SEQUENCE OF 314-339 FROM N.A.  
 RX MEDLINE; 91365255.  
 RA FRIEMERT C., CLOSS E.I., SILBERMANN M., EFFLE V., STRAUSS P.G.;  
 RL GENE 103:259-261(1991).  
 CC -1- FUNCTION: THIOL PROTEASE WHICH IS BELIEVED TO PARTICIPATE IN  
 INTRACELLULAR DEGRADATION AND TURNOVER OF PROTEINS. HAS ALSO  
 BEEN IMPLICATED IN TUMOR INVASION AND METASTASIS.  
 CC -1- CATALYTIC ACTIVITY: HYDROLASES PROTEINS, WITH A SPECIFICITY  
 RESEMBLING THAT OF PAPAIN.  
 CC -1- SUBUNIT: DIMER OF A HEAVY CHAIN AND A LIGHT CHAIN CROSS-LINKED  
 BY A DISULFIDE BOND.  
 CC -1- SUBCELLULAR LOCATION: LYSOSOMAL.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE  
 PAPAIN FAMILY OF THIOL PROTEASES.  
 DR EMBL: M65270; G309152; -.  
 DR EMBL: M65263; G309152; JOINED.  
 DR EMBL: M65264; G309152; JOINED.  
 DR EMBL: M65265; G309152; JOINED.  
 DR EMBL: M65266; G309152; JOINED.  
 DR EMBL: M65267; G309152; JOINED.  
 DR EMBL: M65268; G309152; JOINED.  
 DR EMBL: M65269; G309152; JOINED.  
 DR EMBL: M14222; G309202; -.  
 DR EMBL: X54966; G50597; -.  
 DR EMBL: S69034; G239907; -.  
 DR PIR: B26498; KHMSE.  
 DR PIR: S12901; S12901.  
 DR PIR: A38458; A38458.  
 DR PIR: A49826; A49826.  
 DR HSSP: P07858; IHUC.  
 DR MGD: MGI:88561; CTSE.  
 DR PROSITE: PS00139; THIOL\_PROTEASE\_CYS; 1.  
 DR PROSITE: PS00639; THIOL\_PROTEASE\_HIS; 1.  
 DR PROSITE: PS00640; THIOL\_PROTEASE\_ASN; 1.  
 DR KMW: KMW:18561; THIOL\_PROTEASE; LYSOSOME; GLYCOPROTEIN; ZMOGEN; SIGNAL.  
 FT SIGNAL 1 17  
 FT PROPEP 18 79  
 FT CHAIN 80 333  
 FT CHAIN 80 126  
 FT CHAIN 129 333  
 FT CHAIN 334 339  
 FT ACT\_SITE 108 108  
 FT ACT\_SITE 278 278  
 FT ACT\_SITE 298 298  
 FT DISULFID 93 122  
 FT DISULFID 105 150  
 FT DISULFID 141 207  
 FT DISULFID 142 146  
 FT DISULFID 179 211  
 FT DISULFID 187 198  
 FT CARBOHYD 192 192  
 FT CONFLICT 160 160  
 FT CONFLICT 174 174  
 FT CONFLICT 177 177  
 FT CONFLICT 284 284  
 SQ SEQUENCE 339 AA; 37280 MW; 30694B74 CRC32;

Query Match 85.4%; Score 41; DB 1; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 1.23e+01;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 60 GTVLGG 65



\*\*\*\*\*  
MUSE (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Oct 28 11:30:57 1998; Maspar time 2.19 Seconds  
Tabular output not generated. 80.154 Million cell updates/sec

Title: >US-08-968-685-11  
Description: (1-7) from US08968685.pep  
Perfect Score: 48  
Sequence: 1 GTVLGK 7

Scoring table: PAM 150  
Gap 15

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot35  
1:swiss1

Statistics: Mean 21.225; Variance 20.046; scale 1.059

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	43	89.6	488	1	DSBD_ECOLI THIOI:DISULFIDE INTERC	3.96e+00
2	42	87.5	1029	1	KSP1_YEAST SERINE/THREONINE-PROTE	7.01e+00
3	41	85.4	148	1	YORX_PYRMO HYPOTHETICAL PROTEIN I	1.23e+01
4	41	85.4	299	1	YC21_METUA HYPOTHETICAL PROTEIN M	1.23e+01
5	41	85.4	301	1	CUT2_SCHPO CUT2 PROTEIN.	1.23e+01
6	41	85.4	339	1	CATB_MOUSE CATHEPSIN B PRECURSOR	1.23e+01
7	41	85.4	339	1	CATB_MOUSE CATHEPSIN B PRECURSOR	1.23e+01
8	41	85.4	475	1	Y294_MYCPN HYPOTHETICAL PROTEIN M	1.23e+01
9	41	85.4	507	1	FLGR_BACSU FLAGELLAR HOOK-ASSOCIA	1.23e+01
10	41	85.4	1144	1	CY43_RAT ADENYLATE CYCLASE. OLF	1.23e+01
11	39	81.3	117	1	LY0A_HUMAN IG LAMBDA CHAIN V REGI	3.65e+01
12	39	81.3	275	1	TRY4_ANOGA TRYPSIN 4 PRECURSOR (E	3.65e+01
13	39	81.3	299	1	ISPA_ECOLI GERANYLTRANSTRANSFERAS	3.65e+01
14	39	81.3	321	1	MRKD_KLEPN FIMBRIN ADHESIN PROTEI	3.65e+01
15	39	81.3	361	1	OXDA_FUSPO D-AMINO ACID OXIDASE (	3.65e+01
16	39	81.3	377	1	PRGR_SHEEP PROGESTERONE RECEPTOR	3.65e+01
17	39	81.3	472	1	CRTI_SYNY3 PHYTOENE DEHYDROGENASE	3.65e+01
18	39	81.3	474	1	CRTI_SYNP7 PHYTOENE DEHYDROGENASE	3.65e+01
19	39	81.3	521	1	GAG_HV2D2 GAG POLYPROTEIN (CORE	3.65e+01
20	39	81.3	566	1	CRTI_ARATH PHYTOENE DEHYDROGENASE	3.65e+01
21	39	81.3	570	1	CRTI_SOYBN PHYTOENE DEHYDROGENASE	3.65e+01
22	39	81.3	571	1	CRTI_MAIZE PHYTOENE DEHYDROGENASE	3.65e+01
23	39	81.3	578	1	AC22_STRCO PROBABLE ACTINORHODIN	3.65e+01

RESULT	1	STANDARD;	PRT;	488 AA.
ID	DSBD_ECOLI			
AC	P36655; P76796;			
DT	01-JUN-1994 (REL. 29, CREATED)			
DT	01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	THIOI:DISULFIDE INTERCHANGE PROTEIN DSBD (C-TYPE CYTOCHROME			
DE	BIOGENESIS PROTEIN CYC2) (INNER MEMBRANE COPPER TOLERANCE PROTEIN).			
GN	DSBD OR CYC2 OR CUTA2 OR DIPZ.			
OS	ESCHERICHIA COLI			
OC	PROKARYOTA, GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;			
OC	ENTEROBACTERIACEAE.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-K12;			
RX	MEDLINE: 95349398.			
RA	CROOKS H., COLE J.			
RL	MOL. MICROBIOL. 15:1139-1150(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-K12 / W3110.			
RX	MEDLINE: 95349397.			
RA	FONG S.-T., CAMAKARIS J., LEE B.T.O.;			
RL	MOL. MICROBIOL. 15:1127-1137(1995).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-K12 / MG1655;			
RX	MEDLINE: 95349362.			
RA	BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,			
RL	BLATTNER F.R.;			
RN	NUCLEIC ACIDS RES. 23:2105-2119(1995).			
RN	[4]			
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.			
RC	MEDLINE: 95354659.			
RA	MISTAKAS D., SCHWAGER F., RAINA S.;			
RL	EMBO J. 14:3415-3424(1995).			
RN	[5]			
RP	SEQUENCE OF 1-5.			
RC	MISTAKAS D., HUGHES G.J., FRUTIGER S., PAQUET N., RAINA S.;			
RL	SUBMITTED (MAY-1995) TO THE SWISS-PROT DATA BANK.			
RN	[6]			
RP	CHARACTERIZATION.			
RC	MEDLINE: 95046265.			
RA	SAMBONGI Y., FERGUSON S.J.;			
RL	FERS LETT. 353:235-238(1994).			
CC	-1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME			
CC	PERIPLASMIC PROTEINS. ONE OF ITS FUNCTION COULD BE TO MAINTAIN			



B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, E.; Vassart, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weizenecker, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

#journal  
#title The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

#accession H69622

#status preliminary; nucleic acid sequence not shown;

#molecule-type DNA

#residues 1-507 #label KUN

#experimental-source strain 168

GENETICS

#gene flgK

SUMMARY

#length 507 #molecular-weight 54355 #checksum 3751

Query Match 85.4%; Score 41; DB 2; Length 507;  
Best Local Similarity 85.7%; Pred. No. 4.64e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 256 GTVLDGK 262

QY 1 GTVLGK 7

RESULT 13

ENTRY A39833 #type complete

TITLE adenylyl cyclase (EC 4.6.1.1) type III - rat

ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat

DATE 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 10-Sep-1997

ACCESSIONS A39833

REFERENCE A39833

#authors Bakalyar, H.A.; Reed, R.R.

#journal Science (1990) 250:1403-1406

#title Identification of a specialized adenylyl cyclase that may mediate odorant detection.

#cross-references MUID:91075227

#accession A39833

#molecule-type mRNA

#residues 1-1144 #label BAK

CLASSIFICATION ##cross-references GB:M55075; NID:g202714; PID:g202715

KEYWORDS #superfamily guanylate cyclase catalytic domain homology  
glycoprotein; lyase; olfaction; phosphorus-oxygen lyase;  
transmembrane protein

FEATURE

258-496 #domain guanylate cyclase catalytic domain homology

871-1123 #label GCC1  
#domain guanylate cyclase catalytic domain homology  
#label GCC2  
#binding\_site carbohydrate (asn) (covalent) #status predicted

SUMMARY

#length 1144 #molecular-weight 128935 #checksum 17

Query Match 85.4%; Score 41; DB 2; Length 1144;  
Best Local Similarity 100.0%; Pred. No. 4.64e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 418 GTVLGG 423

QY 1 GTVLGG 6

RESULT 14

ENTRY S58775 #type complete

TITLE mypI protein - smut fungus (*Ustilago maydis*)

ORGANISM #formal\_name Ustilago maydis #common\_name corn smut

DATE 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change

20-Feb-1998

ACCESSIONS S58775

REFERENCE S58775

#authors Glasson, L.; Kronstad, J.W.

#journal Genetics (1995) 141:491-501

#title Mutations in the mypI gene of *Ustilago maydis* attenuate mycelial growth and virulence.

#accession S58775

#status preliminary

#molecule-type DNA

#residues 1-1150 #label G1A

#cross-references EMBL:133919

GENETICS

#gene mypI

SUMMARY

#length 1150 #molecular-weight 123387 #checksum 61

Query Match 85.4%; Score 41; DB 2; Length 1150;  
Best Local Similarity 100.0%; Pred. No. 4.64e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1019 GTVLGG 1024

QY 1 GTVLGG 6

RESULT 15

ENTRY A70145 #type complete

TITLE prolipoprotein diacylglycerol transferase (lgt) homolog - Lyme disease spirochete

ORGANISM #formal\_name Borrelia burgdorferi #common\_name Lyme disease spirochete

DATE 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 13-Feb-1998

ACCESSIONS A70145

REFERENCE A70100

#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.R.; Gwinn, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kierlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Uteirack, T.; Matthey, L.; McDonald, L.; Artach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.

#journal Nature (1997) 390:580-586

#title Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.

#cross-references MUID:98065943

#accession A70145

#status preliminary; nucleic acid sequence not shown;

#molecule-type DNA

#residues 1-328 #label K1E

CLASSIFICATION ##cross-references GB:A2000783; TIGR:BB0362

KEYWORDS #experimental-source strain B31

SUMMARY

#length 328 #molecular-weight 37595 #checksum 2651

Query Match 83.3%; Score 40; DB 2; Length 328;  
Best Local Similarity 57.1%; Pred. No. 7.37e+01;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 65 GAILGR 71

QY 1 GTVLGG 7

Search completed: Wed Oct 28 11:32:49 1998

Job time : 27 secs.



Db 34 GTVLGG 39  
111111  
QY 1 GTVLGG 6

RESULT 8  
ENTRY 1MTRB #type complete  
TITLE procathepsin b (EC 3.4.22.1) mutant C29S, S115A, chain B -  
PDB TITLE rat  
ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
#note var. naturally occurring variant Y233a, expressed in pichia  
pastoris

REFERENCE  
#authors A66136  
Cyler, M.; Sivaraman, J.; Grochulski, P.; Coulombe, R.;  
Storer, A.C.; Mort, J.S.

#submission submitted to the Brookhaven Protein Data Bank, January 1996  
#cross-references PDB:1MTR  
#comment Resolution: 2.8 angstroms  
#comment Determination: X-ray diffraction  
#comment R-value: no refinement  
#keywords cysteine protease; hydrolase; thiol protease

FEATURE  
2-11 #region helix (right hand alpha)  
26-32 #region helix (right hand alpha)  
60-1 #region helix (right hand alpha)  
6-9 #region helix (right hand alpha)  
20-37 #region helix (right hand alpha)  
47-53 #region helix (right hand alpha)  
117-119 #region helix (right hand 3-10)  
123-125 #region helix (right hand 3-10)  
129-138 #region helix (right hand alpha)  
193-195 #region helix (right hand alpha)  
210-220 #region helix (right hand alpha)  
231-235 #region helix (right hand alpha)  
292-294 #region helix (right hand 3-10)  
307-309 #region helix (right hand 3-10)  
259-262, 265-271, #region beta sheet  
283-287 #region beta sheet  
15-17, 236-239 #region beta sheet  
200-206, 299-304 #region beta sheet  
223-230, 248-256 #region beta sheet  
5-34 #disulfide\_bonds  
17-124 #disulfide\_bonds  
53-181 #disulfide\_bonds  
116-120 #disulfide\_bonds  
153-185 #disulfide\_bonds  
161-172 #disulfide\_bonds  
20, 252, 272 #site Ser, His, Asn #label CVB

SUMMARY  
#length 313 #molecular-weight 34537 #checksum 1518

Query Match 85.4%; Score 41; DB 5; Length 313;  
Best Local Similarity 100.0%; Pred. No. 4.64e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 34 GTVLGG 39  
111111  
QY 1 GTVLGG 6

RESULT 9  
ENTRY KHRB #type complete  
TITLE cathepsin B (EC 3.4.22.1) precursor - rat  
ALTERNATE\_NAMES cathepsin B1; RSG-2 protein  
ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
#note 18-Apr-1984 #sequence\_revision 01-Dec-1995 #text\_change  
18-Mar-1997

ACCESSIONS  
#accession S51041, A00977, I59019  
#molecule\_type DNA  
#residues 1-339 #label Q1A

REFERENCE  
#authors Guenette, R.S.; Woolbroek, M.; Wong, K.; Wong, P.;  
Tenniswood, M.  
Eur. J. Biochem. (1994) 226:311-321  
#journal Cathepsin B, a cysteine protease implicated in metastatic

progression, is also expressed during regression of the rat  
prostate and mammary glands.

#accession S51041  
#status Preliminary  
#molecule\_type mRNA  
#residues 1-339 #label GUE

REFERENCE  
#authors A00976  
Takio, K.; Towatari, T.; Katunuma, N.; Teller, D.C.; Titani,  
K.

#journal Proc. Natl. Acad. Sci. U.S.A. (1983) 80:3666-3670  
#title Homology of amino acid sequences of rat liver cathepsins B  
and H with that of papain.

#cross-references MUID:83221657  
#accession A00977  
#molecule\_type protein  
#residues 80-126, 129-158, 'G', 160-333 #label TAK

REFERENCE  
#authors I59019  
San Segundo, B.; Chan, S.J.; Steiner, D.F.  
Proc. Natl. Acad. Sci. U.S.A. (1985) 82:2320-2324  
#title Identification of cDNA clones encoding a precursor of rat  
liver cathepsin B.

#cross-references MUID:85190489  
#accession I59019  
#status translated from GB/EMBL/DBJ  
#molecule\_type mRNA  
#residues 69-339 #label RES  
#cross-references GB:M11305; MUID:9203647; PTD:9203648

FUNCTION  
#description broad specificity endopeptidase and peptidyl dipeptidase  
#pathway intracellular protein degradation  
CLASSIFICATION  
#superfamily papain  
KEYWORDS  
cysteine proteinase; glycoprotein; hydrolase; lysosome;  
protein degradation

FEATURE  
93-122, 105-150, #disulfide\_bonds #status predicted  
141-207, 142-146, #active\_site Cys, His, Asn #status predicted  
179-211, 187-198  
108, 278, 298  
192 #binding\_site carbohydrate (asn) (covalent) #status  
experimental

SUMMARY  
#length 339 #molecular-weight 37470 #checksum 3297

Query Match 85.4%; Score 41; DB 1; Length 339;  
Best Local Similarity 100.0%; Pred. No. 4.64e+01;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 60 GTVLGG 65  
111111  
QY 1 GTVLGG 6

RESULT 10  
ENTRY KMSB #type complete  
TITLE cathepsin B (EC 3.4.22.1) precursor - mouse  
ALTERNATE\_NAMES preprocathepsin B  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
#note 31-Mar-1992 #sequence\_revision 26-Apr-1996 #text\_change  
05-Sep-1997

ACCESSIONS  
#accession A38458, B26498; S12901; PS0360  
#molecule\_type DNA  
#residues 1-339 #label Q1A

REFERENCE  
#authors A49826  
Qian, F.; Frankfater, A.; Chan, S.J.; Steiner, D.F.  
DNA Cell Biol. (1991) 10:159-168  
#title The structure of the mouse cathepsin B gene and its putative  
promoter.

REFERENCE  
#accession A49826  
#cross-references GB:M63999  
#molecule\_type DNA  
#residues 1-339 #label Q1A

REFERENCE  
#authors Qian, F.; Frankfater, A.; Steiner, D.F.; Bajkowski, A.S.;  
Chan, S.J.  
Anticancer Res. (1991) 11:1445-1451  
#journal

```

#authors      #authors
#journal      #journal
#title        #title
#cross-references MUID:903030536
#accession    S10655
##molecule_type DNA
##residues    1-148 ##label 2M1
##note        the authors translated the codon GGT for residue 54 as
                glu
SUMMARY      #length 148 #checksum 8712

Query Match      85.4%; Score 41; DB 2; Length 148;
Best Local Similarity 100.0%; Pred. No. 4.64e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      89 GTVGLG 94
        |||||
Oy      1 GTVGLG 6

RESULT      5
ENTRY      D64452 #type complete
TITLE      hypothetical protein MJ1221 - Methanococcus jannaschii
ORGANISM    #formal_name Methanococcus jannaschii
DATE        13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
                10-Oct-1997
ACCESSIONS  D64452
REFERENCE    A64300
#authors    Butl, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
                R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,
                R.A.; Gocayne, J.D.; Kervilavage, A.R.; Dougherty, B.A.;
                Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
                Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
                Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann,
                J.L.; Nguyen, D.; Uitterlbeck, T.R.; Kelley, J.M.; Peterson,
                J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
                K.M.; Hrust, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
                H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
                Science (1996) 273:1058-1073
                Complete genome sequence of the methanogenic archaeon,
                Methanococcus jannaschii.
#cross-references MUID:96337999
#accession   D64452
#status      preliminary; nucleic acid sequence not shown;
                translation not shown
##molecule_type DNA
##residues    1-299 ##label BUTL
##cross-references GB:067563; GB:D77117; NID:91591850; PID:91591852;
                TIGR:MJ1221; PID:91511223
SUMMARY      #map_position FOR1164064-1164963
                #start_codon TTG
                #length 299 #molecular_weight 32538 #checksum 2920

Query Match      85.4%; Score 41; DB 2; Length 299;
Best Local Similarity 85.7%; Pred. No. 4.64e+01;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db      161 GTVGLG 167
        |||||
Oy      1 GTVGLG 7

RESULT      6
ENTRY      B35694 #type complete
TITLE      cut2 protein - fission yeast (Schizosaccharomyces pombe)
ORGANISM    #formal_name Schizosaccharomyces pombe
DATE        30-Nov-1990 #sequence_revision 30-Nov-1990 #text_change
                09-Sep-1997
ACCESSIONS  B35694

```

```

REFERENCE          A35694
#authors           Uzawa, S.; Samejima,I.; Hirano, T.; Tanaka, K.; Yanagida, M
#journal           Cell (1990) 62:913-925
#title             The fission yeast cut1+ gene regulates spindle pole body
                  duplication and has homology to the budding yeast Esp1
                  gene.
#cross-references  MU1D:90367123
#accession         B35694
##status          preliminary
##molecule-type  DNA
##residues        1-301 #label UZA
#cross-references  GB:M57750; NID:9173374; PID:9173375
SUMMARY            #length 301 #molecular-weight 32854 #checksum 3948

Query Match      85.4%: Score 41; DB 2; Length 301;
Best Local Similarity 100.0%; Pred. No. 4.64e+01;
Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db              53 TVLGGK 58
QY              2 TVLGGK 7

RESULT          7
ENTRY           1MIRA
TITLE           procatepsin b (EC 3.4.22.1) mutant C29S, S115A, chain A -
PDB-TITLE       rat procatepsin b
ORGANISM        rat
#note           #format_name Rattus norvegicus #common_name Norway rat
                  var. naturally occurring variant V223a, expressed in plichia
                  pastoris
REFERENCE        A66136
#authors         Cygler, M.; Sivaraman, J.; Grochulski, P.; Coulombe, R.;
                  Scorer, A.C.; Mott, J.S.
#submission      submitted to the Brookhaven Protein Data Bank, January 1996
#cross-references PDB:1MIR
COMMENT          Resolution: 2.8 angstroms
COMMENT          Determination: X-ray diffraction
COMMENT          R-value: no refinement
KEYWORDS         cysteine protease; hydrolase; thiol protease
FEATURE          2-11
                26-32
                60-1
                6-9
                20-37
                47-53
                117-119
                123-125
                129-138
                193-195
                210-220
                231-235
                292-304
                307-309
                259--262,265-271,
                283-287
                15-17,236-239
                200-206,298-304
                223-230,246-256
                5-34
                17-124
                53-181
                116-120
                153-185
                161-172
                20,252,272
                #length 313 #molecular-weight 34537 #checksum 1518

SUMMARY          Query Match      85.4%: Score 41; DB 5; Length 313;
                  Best Local Similarity 100.0%; Pred. No. 4.64e+01;
                  Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

ALTERNATE\_NAMES thiol:disulfide interchange protein dsbd  
ORGANISM #formal\_name Escherichia coli  
DATE 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change  
21-Nov-1997  
ACCESSIONS S56364; I41028; I41037; S42064; F65223; S57220; S47295  
REFERENCE  
#authors Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.;  
Blattner, F.R.  
#journal Nucleic Acids Res. (1995) 23:2105-2119  
#title Analysis of the Escherichia coli genome VI: DNA sequence of  
the region from 92.8 through 100 minutes.  
#cross-references M01D:9534362  
#accession S56364  
#status nucleic acid sequence not shown; translation not shown  
#molecule\_type DNA  
#residues 1-565 #label BUR  
#cross-references EMBL:U14003; NID:g1263172; PID:g536980  
#note The nucleotide sequence was submitted to the EMBL data  
Library, August 1994  
REFERENCE  
#authors Fong, S.T.; Camakaris, J.; Lee, B.T.  
#journal Mol. Microbiol. (1995) 15:1127-1137  
#title Molecular genetics of a chromosomal locus involved in copper  
tolerance in Escherichia coli K-12.  
#cross-references M01D:95349397  
#accession I41028  
#status translated from GB/EMBL/DBJ  
#molecule\_type DNA  
#residues 'M', 78-565 #label RES  
#cross-references EMBL:Z36905; NID:g555290; PID:g581055  
#note In this report, the codon GTG for Val-77 was interpreted  
as a start codon  
REFERENCE  
#authors Crooke, H.; Cole, J.  
#journal Mol. Microbiol. (1995) 15:1139-1150  
#title The biogenesis of C-type cytochromes in Escherichia coli  
requires a membrane-bound protein, DipZ, with a protein  
disulphide isomerase-like domain.  
#cross-references M01D:95349398  
#accession I41037  
#status translated from GB/EMBL/DBJ  
#molecule\_type DNA  
#residues 'M', 78-565 #label RE2  
#cross-references EMBL:X77707; NID:g871027; PID:g871029  
#note In this report, the codon GTG for Val-77 was interpreted  
as a start codon  
REFERENCE  
#accession S42064  
#molecule\_type DNA  
#residues 1-328, 'V', 330-565 #label CRO  
#cross-references EMBL:X77707  
REFERENCE  
#authors Blatter, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;  
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;  
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;  
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,  
Y.  
#journal Science (1997) 277:1453-1462  
#title The complete genome sequence of Escherichia coli K-12.  
#cross-references M01D:97426617  
#accession F65223  
#status nucleic acid sequence not shown; translation not shown  
#molecule\_type DNA  
#residues 1-365 #label BLAT  
#cross-references GB:AE000486; GB:U00096; NID:g1790574; PID:g1790578;  
UMG:B4136  
#experimental\_source strain K-12, substrain MG1655  
REFERENCE S57220

#authors Mistakias, D.; Schwager, F.; Raine, S.  
#journal EMBL J. (1995) 14:3415-3424  
#title Identification and characterization of a new disulfide  
isomerase-like protein (DsbD) in Escherichia coli.  
#accession S57220  
#status Preliminary  
#molecule\_type DNA  
#residues 462-540 #label MIS  
GENETICS  
#gene dsbD; cycZ; CutA2; dipZ  
#keywords inner membrane; redox-active disulfide; transmembrane protein  
SUMMARY #length 565 #molecular\_weight 61795 #checksum 2685  
Query Match 89.6%; Score 43; DB 2; Length 565;  
Best Local Similarity 85.7%; Pred. No. 1.79e+01;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Db 192 GTVLGK 198  
QY 1 GTVLGK 7  
RESULT 3  
ENTRY S64731 #type complete  
TITLE serine/threonine-specific kinase KSP1 (EC 2.7.1.-) - yeast  
ALTERNATE\_NAMES (Saccharomyces cerevisiae)  
ORGANISM hypothetical protein YHR082c  
DATE 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change  
06-Feb-1998  
ACCESSIONS S64731; S46818; S47912  
REFERENCE  
#authors Fleischmann, M.; Stagliar, I.; Aebi, M.  
#journal Mol. Gen. Genet. (1996) 250:614-625  
#title Allele-specific suppression of a Saccharomyces cerevisiae  
prp20 mutation by overexpression of a nuclear  
serine/threonine protein kinase.  
#accession S64731  
#molecule\_type DNA  
#residues 1-1029 #label FLE  
#cross-references EMBL:X80329; NID:g516232; PID:g516233  
REFERENCE S46794  
#authors Favello, T.  
#submission submitted to the EMBL Data Library, June 1994  
#description The sequence of S. cerevisiae cosmid 9205.  
#accession S46818  
#molecule\_type DNA  
#residues 1-1029 #label FAV  
#cross-references EMBL:U10556; NID:g500825; PID:g500839; MIPS:YHR082C  
GENETICS  
#gene SGD:KSP1  
#map\_position 8R  
#keywords phosphotransferase  
SUMMARY #length 1029 #molecular\_weight 117081 #checksum 5423  
Query Match 87.5%; Score 42; DB 2; Length 1029;  
Best Local Similarity 85.7%; Pred. No. 2.89e+01;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 663 STVLGK 669  
QY 1 GTVLGK 7  
RESULT 4  
ENTRY S10655 #type fragment  
TITLE hypothetical protein X - Pyrococcus woesei (fragment)  
ORGANISM #formal\_name Pyrococcus woesei  
DATE 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change  
18-Jun-1993  
ACCESSIONS S10655  
REFERENCE S10650

\*\*\*\*\*  
MORSE  
(TM)  
\*\*\*\*\*

Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Oct 28 11:32:22 1998; Maspar time 3.06 Seconds  
Tabular output not generated. 83.510 Million cell updates/sec

Title: >US-08-968-685-11  
Description: (1-7) from US08968685.pep  
Perfect Score: 48  
Sequence: 1 GVLGK 7

Scoring table: PAM 150  
Gap 15

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r56  
1:p1r1 2:p1r2 3:p1r3 4:p1r4 5:m13d

Statistics: Mean 20.326; Variance 24.108; scale 0.843

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	43	89.6	230	2	C64658	outer membrane proteol
2	43	89.6	565	2	S56364	inner membrane copper
3	42	87.5	1029	2	S64731	serine/threonine-spec
4	41	85.4	148	2	S10655	hypothetical protein
5	41	85.4	299	2	D64452	hypothetical protein
6	41	85.4	301	2	B35694	cult2 protein
7	41	85.4	313	5	1MIRA	procathepsin b (EC 3.
8	41	85.4	313	5	1MIRB	procathepsin b (EC 3.
9	41	85.4	339	1	KHRTB	cathepsin B (EC 3.4.2
10	41	85.4	339	1	KHMSB	cathepsin B (EC 3.4.2
11	41	85.4	475	2	S73746	MG294 homolog A05.crf
12	41	85.4	507	2	H69622	flagellar hook-associ
13	41	85.4	1144	2	A39833	adenylate cyclase (EC
14	41	85.4	1150	2	S58775	mypl protein - smut f
15	40	83.3	328	2	A70145	prolipo protein diacyl
16	39	81.3	97	2	S36066	ig lambda chain - hum
17	39	81.3	98	2	S36062	ig lambda chain - hum
18	39	81.3	112	2	S26655	ig lambda chain V reg
19	39	81.3	117	1	LVH02	ig lambda chain precu
20	39	81.3	117	2	S04526	ig lambda chain precu
21	39	81.3	275	2	S40005	trypsin (EC 3.4.21.4)
22	39	81.3	299	2	J00665	geranyltransferase
23	39	81.3	321	2	B32801	fimbrial adhesin prec

24	39	81.3	361	3	JX0152	D-amino-acid oxidase	1.16e+02
25 <td>39</td> <td>81.3<td>467</td><td>2</td><td>S20939</td><td>phytoene dehydrogenase</td><td>1.16e+02</td></td>	39	81.3 <td>467</td> <td>2</td> <td>S20939</td> <td>phytoene dehydrogenase</td> <td>1.16e+02</td>	467	2	S20939	phytoene dehydrogenase	1.16e+02
26 <td>39</td> <td>81.3<td>472</td><td>2</td><td>S74886</td><td>phytoene desaturase -</td><td>1.16e+02</td></td>	39	81.3 <td>472</td> <td>2</td> <td>S74886</td> <td>phytoene desaturase -</td> <td>1.16e+02</td>	472	2	S74886	phytoene desaturase -	1.16e+02
27 <td>39</td> <td>81.3<td>474</td><td>2</td><td>S16250</td><td>phytoene dehydrogenase</td><td>1.16e+02</td></td>	39	81.3 <td>474</td> <td>2</td> <td>S16250</td> <td>phytoene dehydrogenase</td> <td>1.16e+02</td>	474	2	S16250	phytoene dehydrogenase	1.16e+02
28 <td>39</td> <td>81.3<td>521</td><td>2</td><td>S08435</td><td>gag polypeptide - hum</td><td>1.16e+02</td></td>	39	81.3 <td>521</td> <td>2</td> <td>S08435</td> <td>gag polypeptide - hum</td> <td>1.16e+02</td>	521	2	S08435	gag polypeptide - hum	1.16e+02
29 <td>39</td> <td>81.3<td>570</td><td>2</td><td>A39597</td><td>phytoene dehydrogenase</td><td>1.16e+02</td></td>	39	81.3 <td>570</td> <td>2</td> <td>A39597</td> <td>phytoene dehydrogenase</td> <td>1.16e+02</td>	570	2	A39597	phytoene dehydrogenase	1.16e+02
30 <td>39</td> <td>81.3<td>570</td><td>2</td><td>S54134</td><td>phytoene dehydrogenase</td><td>1.16e+02</td></td>	39	81.3 <td>570</td> <td>2</td> <td>S54134</td> <td>phytoene dehydrogenase</td> <td>1.16e+02</td>	570	2	S54134	phytoene dehydrogenase	1.16e+02
31 <td>39</td> <td>81.3<td>571</td><td>2</td><td>S65060</td><td>phytoene desaturase p</td><td>1.16e+02</td></td>	39	81.3 <td>571</td> <td>2</td> <td>S65060</td> <td>phytoene desaturase p</td> <td>1.16e+02</td>	571	2	S65060	phytoene desaturase p	1.16e+02
32 <td>39</td> <td>81.3<td>582</td><td>2</td><td>B40046</td><td>tetracycline resistanc</td><td>1.16e+02</td></td>	39	81.3 <td>582</td> <td>2</td> <td>B40046</td> <td>tetracycline resistanc</td> <td>1.16e+02</td>	582	2	B40046	tetracycline resistanc	1.16e+02
33 <td>39</td> <td>81.3<td>582</td><td>2</td><td>S29314</td><td>phytoene dehydrogenase</td><td>1.16e+02</td></td>	39	81.3 <td>582</td> <td>2</td> <td>S29314</td> <td>phytoene dehydrogenase</td> <td>1.16e+02</td>	582	2	S29314	phytoene dehydrogenase	1.16e+02
34 <td>39</td> <td>81.3<td>583</td><td>2</td><td>A45381</td><td>phytoene dehydrogenase</td><td>1.16e+02</td></td>	39	81.3 <td>583</td> <td>2</td> <td>A45381</td> <td>phytoene dehydrogenase</td> <td>1.16e+02</td>	583	2	A45381	phytoene dehydrogenase	1.16e+02
35 <td>39</td> <td>81.3<td>671</td><td>2</td><td>S61099</td><td>leukotriene-A4 hydrol</td><td>1.16e+02</td></td>	39	81.3 <td>671</td> <td>2</td> <td>S61099</td> <td>leukotriene-A4 hydrol</td> <td>1.16e+02</td>	671	2	S61099	leukotriene-A4 hydrol	1.16e+02
36 <td>39</td> <td>81.3<td>786</td><td>2</td><td>A35466</td><td>progestosterone receptor</td><td>1.16e+02</td></td>	39	81.3 <td>786</td> <td>2</td> <td>A35466</td> <td>progestosterone receptor</td> <td>1.16e+02</td>	786	2	A35466	progestosterone receptor	1.16e+02
37 <td>39</td> <td>81.3<td>923</td><td>2</td><td>A39596</td><td>progestosterone receptor</td><td>1.16e+02</td></td>	39	81.3 <td>923</td> <td>2</td> <td>A39596</td> <td>progestosterone receptor</td> <td>1.16e+02</td>	923	2	A39596	progestosterone receptor	1.16e+02
38 <td>39</td> <td>81.3<td>923</td><td>2</td><td>I53280</td><td>progestosterone receptor</td><td>1.16e+02</td></td>	39	81.3 <td>923</td> <td>2</td> <td>I53280</td> <td>progestosterone receptor</td> <td>1.16e+02</td>	923	2	I53280	progestosterone receptor	1.16e+02
39 <td>39</td> <td>81.3<td>930</td><td>2</td><td>A25923</td><td>progestosterone receptor</td><td>1.16e+02</td></td>	39	81.3 <td>930</td> <td>2</td> <td>A25923</td> <td>progestosterone receptor</td> <td>1.16e+02</td>	930	2	A25923	progestosterone receptor	1.16e+02
40 <td>39</td> <td>81.3<td>933</td><td>1</td><td>ORHUP</td><td>progestosterone receptor</td><td>1.16e+02</td></td>	39	81.3 <td>933</td> <td>1</td> <td>ORHUP</td> <td>progestosterone receptor</td> <td>1.16e+02</td>	933	1	ORHUP	progestosterone receptor	1.16e+02
41 <td>39</td> <td>81.3<td>1109</td><td>2</td><td>A40801</td><td>phosphoprotein phosph</td><td>1.16e+02</td></td>	39	81.3 <td>1109</td> <td>2</td> <td>A40801</td> <td>phosphoprotein phosph</td> <td>1.16e+02</td>	1109	2	A40801	phosphoprotein phosph	1.16e+02
42 <td>38</td> <td>79.2</td> <td>325</td> <td>2</td> <td>S51569</td> <td>moet protein - Rhlzob</td> <td>1.81e+02</td>	38	79.2	325	2	S51569	moet protein - Rhlzob	1.81e+02
43 <td>38</td> <td>79.2</td> <td>341</td> <td>2</td> <td>S37438</td> <td>transposase - Strepto</td> <td>1.81e+02</td>	38	79.2	341	2	S37438	transposase - Strepto	1.81e+02
44 <td>38</td> <td>79.2</td> <td>530</td> <td>2</td> <td>A26638</td> <td>homeotic protein Dfd</td> <td>1.81e+02</td>	38	79.2	530	2	A26638	homeotic protein Dfd	1.81e+02
45 <td>38</td> <td>79.2</td> <td>687</td> <td>2</td> <td>H64665</td> <td>carbon starvation pro</td> <td>1.81e+02</td>	38	79.2	687	2	H64665	carbon starvation pro	1.81e+02

## ALIGNMENTS

RESULT	1	ENTRY	1	ENTRY	1	ENTRY	1	ENTRY	1
ENTRY	C64658	#type complete							
TITLE	outer membrane protein - Helicobacter pylori (strain 26695)								
ORGANISM	#formal name Helicobacter pylori								
DATE	09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 10-Oct-1997								

ACCESSIONS  
REFERENCE  
#authors

Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;  
Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,  
H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,  
J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;  
Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.;  
McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;  
Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Uitterback, T.R.;  
Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;  
Fujii, C.; Bowman, C.; Wathey, L.; Wallin, E.; Hayes,  
W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,  
C.M.; Venter, J.C.  
Nature (1997) 388:539-547  
The complete genome sequence of the gastric pathogen  
Helicobacter pylori.

#journal  
#title  
#cross-references  
#accession  
#status  
#molecule\_type  
#residues  
#cross-references  
#GENETICS  
#start\_codon  
#length  
#molecular\_weight  
#checksum  
#SUMMARY

Query Match  
Best Local Similarity 89.6%; DB 2; Length 230;  
Matches 6; Conservative 0; Mismatches 1; Gaps 0;

Db 82 GVLGK 88  
QY 1 GVLGK 7  
RESULT 2  
ENTRY S56364 #type complete  
TITLE inner membrane copper tolerance protein cyc2 - Escherichia coli

CC which a red colour is desired.  
SQ Sequence 582 AA;

Query Match 81.3%; Score 39; DB 19; Length 582;  
Best Local Similarity 71.4%; Pred. No. 2.53e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 549 gavlsgk 555  
1 GTVLGSK 7  
QY

RESULT 15  
ID W01104 standard; Protein; 582 AA.  
AC W01104;  
DT 20-NOV-1996 (first entry)  
DE N. benthamiana phytoene desaturase.  
KW phytoene desaturase; Nicotiana; zeta-carotene; carotenoid;  
KW biosynthesis.  
OS Nicotiana benthamiana.  
PN US539093-A.  
PD 23-JUL-1996.  
RF 16-JUN-1994; 261086.  
RL 16-JUN-1994; US-261086.  
RT (DELL/) DELTA-CIOPPA G R.  
RA (FITZ/) FITZMAURICE W P.  
PA (GRILL/) GRILL L K.  
PA (HELL/) HELLMANN G M.  
PA (KUMAI/) KUMAGAI M H.  
PI DELLA-CIOPPA GR, FITZMAURICE WP, GRILL LK, HELLMANN GM;  
PI Kumagai MH;  
DR WPI; 96-353678/35.  
DR N-PSDB; T40700.  
PT Phytoene desaturase nucleic acids obtd. from Nicotiana species -  
PT used for the biosynthesis or in vitro synthesis of zeta-carotene.  
PS Claim 2: Column 13-24; 24pp: English.  
CC The present sequence is that of a phytoene desaturase (PD) isolated from  
CC Nicotiana benthamiana. The PD is used for the biosynthesis of zeta-  
CC carotene which is useful in the carotenoid biosynthesis pathway.  
CC Carotenoids such as phytoene have been found to be useful in absorbing  
CC ultraviolet radiation, and lycopene has use as a colouring agent in  
CC situations in which a red colour is desired.  
SQ Sequence 582 AA;

Query Match 81.3%; Score 39; DB 19; Length 582;  
Best Local Similarity 71.4%; Pred. No. 2.53e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 549 gavlsgk 555  
1 GTVLGSK 7  
QY

Search completed: Wed Oct 28 11:33:28 1998  
Job time : 21 secs.



AC R04064;  
DT 03-SEP-1990 (first entry)  
DE D-amino acid oxidase gene product.  
KW D-amino acid oxidase; cephalosporin; cephem; E.coli.  
OS Fusarium solani.  
PN EP-364275-A.  
PD 18-APR-1990.  
PF 12-OCT-1989; 310483.  
PR 13-OCT-1988; JP-260332.  
PA (FUTJ) Fujisawa Pharm KK.  
PI Isogai T, Ono H, Kojo H;  
DR WPI: 90-117771/16.  
N-PSDB: Q03971.  
PT D-amino acid oxidase, prodn. -  
by culture of E.coli transformants contg. expression vectors  
originated from Fusarium solani M-0718.  
Disclosure: Fig 7; 38pp; English.  
E.coli transformed to express DAO, which catalyses the enzymatic  
conversion of cephalosporin C to 7-beta-(5-carboxy-5-oxopentamido)cephalosporanic acid (keto-7ACA). 7ACA is an  
important starting point for the production of cephem  
antibiotics.  
SQ Sequence 361 AA;

Query Match 81.3%; Score 39; DB 3; Length 361;  
Best Local Similarity 83.3%; Pred. No. 2.53e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 251 gtl19g 256  
||:||||  
QY 1 GTVLGG 6

RESULT 12  
ID R13001 standard; Protein: 361 AA.  
AC R13001;  
DT 24-SEP-1991 (first entry)  
DE D-amino oxidase from pCFS315.  
KW Cephalosporin; acylase; enzyme; aminocephem; N-deacylation; DAO.  
PN EP-436555-A.  
PD 10-JUL-1991.  
PF 20-DEC-1990; 313988.  
PR 27-DEC-1989; JP-342113.  
20-JUL-1990; JP-193609.  
(FUTJ) FUJISAWA PHARM KK.  
Isogai T, Fukagawa M, Iwami M, Aramori I, Kojo H;  
WPI: 91-202062/28.  
N-PSDB: Q12579.  
PT Single step fermentative prodn. of 7-amino-3-cephem cpds. - by  
culturing Acromonium chrysogenum transformed with vector encoding  
enzyme involved in N-deacylation  
PS Disclosure: Fig 3-11-(1-2); 86pp; English.  
The vector for producing 7-aminocephem cpd. contains a DNA fragment  
prepared by ligating at least one or more promoter(s) for A.  
chrysogenum to gene(s) for enzyme(s) capable of converting the  
cephalosporin cpd. to the corresp. 7-aminocephem cpd. to each other  
in that order from the upstream side to the downstream side by  
conventional methods. An appropriate selective marker,  
autonomously replication sequence, terminator, translation activating  
sequence and so forth may be inserted into the vector at respective  
desired sites.  
CC The enzyme may be cephalosporin C acylase from P. diminuta V22,  
CC catalysing on step conversion; opt. supplemented by DAO as an enzyme  
CC catalysing two-step conversion. The DAO expression plasmid pCFS315  
CC was isolated from E. coli JM109 (FERM BP-1916).  
SQ Sequence 361 AA;

Query Match 81.3%; Score 39; DB 3; Length 361;  
Best Local Similarity 83.3%; Pred. No. 2.53e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Db 251 gtl19g 256  
||:||||

QY 1 GTVLGG 6

RESULT 13  
ID W01106 standard; Protein: 582 AA.  
AC W01106;  
DT 20-NOV-1996 (first entry)  
DE N. tabacum phytoene desaturase.  
KW Phytoene desaturase; Nicotiana; zeta-carotene; carotenoid;  
KW biosynthesis.  
OS Nicotiana tabacum.  
PN US5539093-A.  
PD 23-JUL-1996.  
PF 16-JUN-1994; 261086.  
PR 16-JUN-1994; US-261086.  
PA (DELL/) DELLA-CIOPPA G R.  
PA (FITZ/) FITZMAURICE W P.  
PA (GRILL/) GRILL L K.  
PA (HELL/) HELLMANN G M.  
PA (KUMA/) KUMAGAI M H.  
PI DELLA-CIOPPA GR, Fitzmaurice WP, Grill LK, Hellmann GM;  
WPI: 96-353878/35.  
N-PSDB: T40702.  
PT Phytoene desaturase nucleic acids obtd. from Nicotiana species -  
used for the biosynthesis or in vitro synthesis of zeta-carotene.  
PS Claim 2; Column 35-40; 24pp; English.  
CC The present sequence is that of a phytoene desaturase (PD) isolated from  
CC Nicotiana tabacum. The PD is used for the biosynthesis of zeta-carotene  
CC which is useful in the carotenoid biosynthesis pathway. Carotenoids such  
CC as phytoene have been found to be useful in absorbing ultraviolet  
CC radiation, and lycopene has use as a colouring agent in situations in  
SQ Sequence 582 AA;

Query Match 81.3%; Score 39; DB 19; Length 582;  
Best Local Similarity 71.4%; Pred. No. 2.53e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 549 gavlsgk 555  
||:||||  
QY 1 GTVLGG 7

RESULT 14  
ID W01105 standard; Protein: 582 AA.  
AC W01105;  
DT 20-NOV-1996 (first entry)  
DE N. tabacum phytoene desaturase.  
KW Phytoene desaturase; Nicotiana; zeta-carotene; carotenoid;  
KW biosynthesis.  
OS Nicotiana tabacum.  
PN US5539093-A.  
PD 23-JUL-1996.  
PF 16-JUN-1994; 261086.  
PR 16-JUN-1994; US-261086.  
PA (DELL/) DELLA-CIOPPA G R.  
PA (FITZ/) FITZMAURICE W P.  
PA (GRILL/) GRILL L K.  
PA (HELL/) HELLMANN G M.  
PA (KUMA/) KUMAGAI M H.  
PI DELLA-CIOPPA GR, Fitzmaurice WP, Grill LK, Hellmann GM;  
WPI: 96-353878/35.  
N-PSDB: T40701.  
PT Phytoene desaturase nucleic acids obtd. from Nicotiana species -  
used for the biosynthesis or in vitro synthesis of zeta-carotene.  
PS Claim 2; Column 27-32; 24pp; English.  
CC The present sequence is that of a phytoene desaturase (PD) isolated from  
CC Nicotiana tabacum. The PD is used for the biosynthesis of zeta-carotene  
CC which is useful in the carotenoid biosynthesis pathway. Carotenoids such  
CC as phytoene have been found to be useful in absorbing ultraviolet  
CC radiation, and lycopene has use as a colouring agent in situations in

Query Match 85.4%; Score 41; DB 17; Length 72;  
Best Local Similarity 85.7%; Pred. No. 1.47e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 45 gtlvlgk 51  
|:|||||  
QY 1 GTVLGK 7

RESULT 8  
ID W52229 standard; Protein; 111 AA.  
AC W52229;  
DT 12-JUN-1998 (first entry)  
DE Antibody LD2-10-VL chain sequence.  
KW Antibody; variable heavy chain; VH chain; variable light chain; VL chain;  
KM Rhesus D antigen; anti-Rhesus D immunoglobulin; HDN; therapy;  
OS Idiopathic thrombocytopenic purpura; haemolytic disease of the newborn.  
PN W09749809-A1.  
PD 31-DEC-1997.  
PE 20-JUN-1997; E03253.  
PR 24-JUN-1996; EP-810421.  
PS (ROT-) ROTKREUZSTIFTUNG ZENT LAB BLUTSPENDE.  
PI Amstutz H, Imboden M, Miescher S, Morell A, Stadler B, Vogel M; WPI: 98-077173/07.  
DR N-PSDB: V19753.  
PT New Rhesus D antigen binding poly:peptide(s) - used to neutralise Rhesus D antigen in therapy, e.g. for treating idiopathic thrombocytopenic purpura  
PT Thrombocytopenic purpura  
PS Claim 1: Fig 9B: 68pp; English.  
CC This sequence is the antibody LD2-10-VL chain sequence, which is a polypeptide of the invention. The polypeptides are capable of forming CC include Rhesus D-specific CDR 1, CDR 2, and CDR 3 regions of pairs of CC variable heavy (VH) and variable light (VL) chain sequences. The CC antibodies are active against the Rhesus D antigen. They can be used for CC treating disorders which would benefit from anti-Rhesus D immunoglobulin, e.g. idiopathic thrombocytopenic purpura. They can also be used for the CC protection of Rhesus negative women before or immediately after the birth CC of a Rhesus positive child to prevent haemolytic disease of the newborn CC (HDN) in subsequent pregnancies. In addition, anti-Rhesus D CC immunoglobulin can be used after mistransfusions of Rhesus positive blood CC to Rhesus negative recipients in order to prevent sensitisation to the CC Rhesus D antigen. The products can also be used as diagnostic reagents.  
SQ Sequence 111 AA;

Query Match 81.3%; Score 39; DB 29; Length 111;  
Best Local Similarity 71.4%; Pred. No. 2.53e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 64 gtlvlgk 70  
|:|||||  
QY 1 GTVLGK 7

RESULT 9  
ID R38883 standard; Protein; 129 AA.  
AC R38883;  
DT 08-FEB-1994 (first entry)  
DE Antibody light chain.  
KW Antibody; Light chain; Heavy chain; lambda; Mu.  
OS Homo sapiens.  
FH Key  
FT binding\_site Location/Qualifiers  
FT 42..55 /label= CDR 1.  
FT 71..77 /label= CDR 2.  
FT binding\_site /label= CDR 3.  
FT 110..118  
FT binding\_site /label= CDR 3.  
FT J05184386-A.  
PN 27-JUN-1993.

PF 27-DEC-1991; 346699.  
PR 27-DEC-1991; JP-346699.  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
FA (SUMO ) SUMITOMO SEIYAKU KK.  
DR WPI: 93-269043/34.  
DR N-PSDB: Q47375.  
PT Modified antibody and a recombinant for its prodn. - includes PT L-chain in antibody and being replaced by L-chain like protein contg. PT specified domain  
PS Claim 3; Figure 3; 13pp; Japanese.  
CC DNA was isolated from human anti-pseudomonas aeruginosa antibody CC producing cell HI223. A HI223 gene library was prepared from the CC human genomic DNA and the Mu and lambda genes were screened. The CC phage clone containing the antibody gene was analysed and the base CC sequence of the HI223 antibody gene was determined. The production CC of the antibody was then increased and the antibody modified by the CC inclusion of an L-chain-like protein. The sequence of the protein CC was then determined.  
SQ Sequence 129 AA;

Query Match 81.3%; Score 39; DB 8; Length 129;  
Best Local Similarity 71.4%; Pred. No. 2.53e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 85 gtlvlgk 91  
|:|||||  
QY 1 GTVLGK 7

RESULT 10  
ID R31531 standard; Protein; 130 AA.  
AC R31531;  
DT 25-MAY-1993 (first entry)  
DE HI223 MAb light chain.  
KW Human; recombinant; light chain; monoclonal; antibody; MAb; HI223;  
KM NanaIwa; microbial infection.  
OS Synthetic.  
FH Key  
FT domain Location/Qualifiers  
FT 43..56 /label= CDR1  
FT domain /label= CDR2  
FT 72..78 /label= CDR2  
FT domain 111..119  
FT domain /label= CDR3  
FN J04360696-A.  
PD 14-DEC-1992.  
PF 06-JUN-1991; 163886.  
PR 06-JUN-1991; JP-163886.  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
FA (SUMO ) SUMITOMO SEIYAKU KK.  
DR WPI: 93-032623/04.  
DR N-PSDB: Q36150.  
PT Recombinant human antibody - produced using new human B PT lymphocyte strain as the host  
PS Claim 5; Fig 3; 27pp; Japanese.  
CC This sequence represents the light chain of a recombinant human CC monoclonal antibody (MAb) designated HI223. This sequence was CC introduced in to a NanaIwa cell strain which was cultured. The CC MAb may be used in a drug for the prevention and treatment of CC diseases caused by microbial infection.  
SQ Sequence 130 AA;

Query Match 81.3%; Score 39; DB 6; Length 130;  
Best Local Similarity 71.4%; Pred. No. 2.53e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 86 gtlvlgk 92  
|:|||||  
QY 1 GTVLGK 7

RESULT 11  
ID R04064 standard; Protein; 361 AA.

ID	Accession	Protein	Score	DB	Length	Pos
RESULT 5						
WT	W22610	standard; Protein; 1891 AA.				
AC	W22610					
DE	27-FEB-1998	(first entry)				
DE	Platenolide synthase ORF5 protein.					
KW	tylactone synthase gene cluster; tylG gene; multifunctional protein;					
KW	polyketide; tylactone synthesis; antibiotic; tylosin.					
OS	Streptomyces ambotaciens.					
FH	Key	Location/Qualifiers				
FT	Domain	35..458				
FT		/note="ketosynthase domain, KS7"				
FT	Domain	582..917				
FT		/note="acyltransferase domain, AT7"				
FT	Domain	1233..1418				
FT		/note="ketoreductase domain, KR7"				
FT	Domain	1502..1585				
FT		/note="acyl carrier protein domain, ACP7"				
FT		1715..1747				
FT		/note="thioesterase domain, TE7"				
PN	EP-791655-A2.					
PD	27-AUG-1997					
PF	19-FEB-1997	301056.				
PR	22-FEB-1996	US-012078.				
PA	(ELI ) LILLY & CO ELI.					
PI	Dehoff BS, Rostock SA, Sutton KL,					
PR	WPI. 97-418046/39.					
DR	N-PEDB: T80414.					
PT	DNA encoding Streptomyces fradiae tylactone synthase domain - for					
PT	production of tylosin-related polyketide compounds					
PS	Example 2; Pages 172-178; 220pp; English.					
CC	W22606-W22610 represent proteins encoded by the platenolide synthase gene					
CC	cluster. The gene cluster is also referred to as the smg gene, and was					
CC	isolated from Streptomyces ambotaciens. These sequences are					
CC	multifunctional proteins which direct the synthesis of the polyketide					
CC	antibiotic platenolide. Platenolide is the basic building block of the macroide					
CC	antibiotic spiramycin. The DNA encoding this sequence was used along with					
CC	the tylG gene (see T80413) to create a hybrid ORF1 sequence (see T80415).					
CC	The tylG gene is the tylactone synthase gene cluster of the invention.					
CC	The tylG sequence was isolated from Streptomyces fradiae, and encodes					
CC	multifunctional proteins which direct the synthesis of the polyketide					
CC	tylactone. Tylactone is the basic building block of the antibiotic					
CC	tylosin. The hybrid sequence can be used to transform S. ambotaciens					
CC	lacking the smg ORF1 sequence, or S. fradiae lacking the tylG ORF1					
CC	sequence, so that they can produce polyketides. The DNA sequence can be					
CC	modified so as to alter the type of carboxylic acids incorporated, the					
CC	number of carboxylic acids incorporated and/or the post-condensation					
CC	reactions performed, thereby resulting in novel tylosin-related					
CC	polyketides.					
SO	Sequence 1891 AA;					
Query Match	87.5%;	Score 42;	DB 26;	Length 1891;		
Best Local Similarity	85.7%;	Pred. No. 1,11e+02;				
Matches	6;	Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
Db	1340 ggv19gk 1346					
Oy	1 GTVLGCK 7					
RESULT 6						
ID	W23720	standard; Protein; 1891 AA.				
AC	W23720					
DE	27-FEB-1998	(first entry)				
DE	Platenolide synthase ORF5 protein.					
KW	Platenolide synthase gene cluster; platenolide production; smg gene;					
KW	multi-functional protein; macroide antibiotic; spiramycin.					
OS	Streptomyces ambotaciens.					
FH	Key	Location/Qualifiers				
FT	Domain	35..458				
FT		/note="ketosynthase domain, KS7"				
FT	Domain	582..917				
FT		/note="acyltransferase domain, AT7"				
FT	Domain	1233..1418				
FT		/note="ketoreductase domain, KR7"				
FT	Domain	1502..1585				
FT		/note="acyl carrier protein domain, ACP7"				
FT		1715..1747				
FT		/note="thioesterase domain, TE7"				
PN	EP-791655-A2.					
PD	27-AUG-1997					
PF	19-FEB-1997	301056.				
PR	22-FEB-1996	US-012078.				
PA	(ELI ) LILLY & CO ELI.					
PI	Dehoff BS, Rostock SA, Sutton KL,					
PR	WPI. 97-418046/39.					
DR	N-PEDB: T80414					

FT	Domain	/note="ketoreductase domain, KR7"
FT	Domain	1502..1585
FT	Domain	/note="acyl carrier protein domain, ACP7"
FT	Domain	1715..1747
FT	Domain	/note="thioesterase domain, TE7"
PN	EP-791656-A2.	
PN	27-AUG-1997.	
PD	19-FEB-1997.	301066.
PF	22-FEB-1996.	US-012050.
PA	(Etil.) Lilly & CO Etil.	
PI	Burgelt SG, Kunstoss SA, Rao RN, Richardson MA;	
PI	Rosteck P;	
DR	WPI: 97-418047/39.	
DR	N-PSDB: T78508.	
PT	DNA encoding Streptomyces ambofaciens platenolide synthase domain -	
PT	for production of spiramycin-related polyketide antibiotics	
PS	Claim 8; Pages 71-77; 81pp; English.	
CC	W23716-W237160 invention. Proteins encoded by the platenolide synthase gene	
CC	cluster of the invention. The gene cluster is also referred to as the	
CC	smg gene, and was isolated from Streptomyces ambofaciens. These	
CC	sequences are multi-functional proteins which direct the synthesis of the	
CC	polyketide platenolide. Platenolide is the basic building block of the	
CC	macrolide antibiotic spiramycin. The DNA can be used to produce compounds	
CC	exhibiting antibiotic activity based on the platenolide structure,	
CC	including specifically the macrolide antibiotic spiramycin and spiramycin DNA	
CC	analogues and derivatives. Modifications of the platenolide synthase DNA	
CC	sequence can be made so as to change the number and type of carboxylic	
CC	acids incorporated into the growing polyketide chain and to change the	
CC	kind of post-condensation processing that is conducted.	
Sequence	1891 AA;	
Query Match	87.5%;	Score 42; DB 26; Length 1891;
Best Local Similarity	85.7%;	Pred. No. 1.11e+02;
Matches	6; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Db	1340 ggv1gk 1346	
Qy	1 GTVLGK 7	
RESULT	7	
ID	R90556 standard; protein; 72 AA.	
AC	R90556:	
DT	08-AUG-1996 (first entry)	
DE	puG4-5-CDK-BP clone derived CDK4 binding protein.	
KW	Cell cycle; CDK4; regulation; G1 phase; proliferation; tumorigenesis;	
KW	cyclin dependent kinase; differentiation; CDK4 inhibitor; agonist;	
OS	antagonist.	
OS	Synthetic.	
PN	W09533818-A2.	
PD	14-DEC-1995.	
PF	02-JUN-1995; US07113.	
PR	02-JUN-1994; US-253155.	
PA	(MITO-) MITOTIX INC.	
PI	Draetta G, Gyuris J;	
DR	WPI: 96-040227/04.	
DR	N-PSDB: T12184.	
PT	Cyclin-dependent kinase-4 binding protein - used in the isolation of	
PT	(ant)agonists of cell cycle regulation.	
PS	Claim 1; Page 92; 115pp; English.	
CC	R90553-R90556 are cyclin dependent kinase 4 (CDK4) binding proteins.	
CC	encoded by clones of the plasmid puG4-5-CDKBP. CDK4 binding proteins	
CC	(CDK4-BP) may be used in an assay for screening test compounds as	
CC	inhibitors of CDK/CDK4-BP interaction. The complexes formed by CDK4	
CC	and D-type cyclins are strongly implicated in the control of the early	
CC	G1 phase of the cell cycle and are strong candidates for controlling	
CC	and/or preventing tumorigenesis and the onset of cancer. Nucleic acids	
CC	encoding CDK4-BP or fragments of these may be used as probes/primers	
CC	to diagnose the presence or absence of genetic lesions in a gene	
CC	encoding 1 of the 24 CDK4-BP, and hence to diagnose the risk for a	
CC	subject of developing a cell-proliferation associated disorder (e.g.	
CC	cancer).	
Sequence	72 AA;	
Qy		

KW Outer membrane protein; OMP; immunogen; vaccine; otitis media; diagnosis.  
 OS Mycobacterium catarrhalis strain 4223.  
 PN MO634960-A1.  
 PD 07-NOV-1996.  
 PF 29-APR-1996; CA0264.  
 PR 01-MAY-1995; US-431718.  
 PR 07-JUN-1995; US-478370.  
 PR 26-MAR-1996; US-621944.  
 PA (CONN-) CONNAUGHT LAB LTD.  
 PI Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K; WPI: 96-506162/50.  
 DR N-PSDB; T38740.  
 PT Moraxella outer membrane protein - useful as immunogen in protective vaccine and for diagnosis.  
 PS Claim 14; Fig 6; 109pp; English.  
 CC An approx. 200 kDa outer membrane protein (MO4505) can be isolated from Moraxella catarrhalis otitis media strain 4223 by electroelution, or expressed from a gene (see also T38740) obtd. from a strain 4223 genomic library. Natural or recombinant outer membrane protein is useful as an immunogen to protect CC against infection by Moraxella, esp. M. catarrhalis. It can also be used to detect antibodies, esp. for differential diagnosis between bacteria that cause similar symptoms, and also useful as a carrier for other antigens and used to raise antitumour antibodies for conjugation to therapeutic agents.  
 SC Sequence 1992 AA;

Query Match 100.0%; Score 48; DB 20; Length 1992;  
 Best Local Similarity 100.0%; Pred. No. 2.05e+01;  
 Matches 7; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 1604 gtvlgk 1610  
 QY 1 GTVLGK 7

RESULT 3  
 ID W23719 standard; Protein; 1580 AA.  
 AC W23719;  
 DT 27-FEB-1998 (first entry)  
 DE Platenolide synthase ORF4 protein.  
 KW Platenolide synthase gene cluster; platenolide production; smg gene;  
 KW multi-functional protein; macrolide antibiotic; spiramycin.  
 OS Streptomycetes ambofaciens.  
 FH Key Location/Qualifiers  
 FT Domain 34..456  
 FT /note="ketosynthase domain, KS6"  
 FT 566..898  
 FT /note="acyltransferase domain, AT6"  
 FT 1148..1333  
 FT /note="ketoreductase domain, KR6"  
 FT 1420..1503  
 FT /note="acyl carrier protein domain, ACP6"  
 Domain  
 EP-791655-A2.  
 PD 27-AUG-1997.  
 PD 19-FEB-1997; 301056.  
 PR 22-FEB-1996; US-012050.  
 PR (ELIL) LILLY & CO ELI.  
 PI Burgett SG, Kuhstoss SA, Rao RN, Richardson MA; Rosteck PR; WPI: 97-418047/39.  
 DR N-PSDB; T78508.  
 PT DNA encoding Streptomycetes ambofaciens platenolide synthase domain - for production of spiramycin-related polyketide antibiotics  
 PS Claim 8; Pages 66-71; 81pp; English.  
 CC W23716-W23720 represent proteins encoded by the platenolide synthase gene cluster of the invention. The gene cluster is also referred to as the smg gene, and was isolated from Streptomycetes ambofaciens. These sequences are multi-functional proteins which direct the synthesis of the polyketide platenolide. Platenolide is the basic building block of the macrolide antibiotic spiramycin. The DNA can be used to produce compounds exhibiting antibiotic activity based on the platenolide structure.

CC Including specifically the macrolide antibiotic spiramycin and spiramycin analogues and derivatives. Modifications of the platenolide synthase DNA sequence can be made so as to change the number and type of carboxylic acids incorporated into the growing polyketide chain and to change the kind of post-condensation processing that is conducted.  
 SC Sequence 1580 AA;

Query Match 87.5%; Score 42; DB 26; Length 1580;  
 Best Local Similarity 85.7%; Pred. No. 1.11e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1255 ggvlgk 1261  
 QY 1 GTVLGK 7

RESULT 4  
 ID W22609 standard; Protein; 1580 AA.  
 AC W22609;  
 DT 27-FEB-1998 (first entry)  
 DE Platenolide synthase ORF4 protein.  
 KW Tyactone synthase gene cluster; tylg gene; multifunctional protein;  
 KW polyketide; tyactone synthesis; antibiotic; tylosin.  
 OS Streptomycetes ambofaciens.  
 FH Key Location/Qualifiers  
 FT Domain 34..456  
 FT /note="ketosynthase domain, KS6"  
 FT 566..898  
 FT /note="acyltransferase domain, AT6"  
 FT 1148..1333  
 FT /note="ketoreductase domain, KR6"  
 FT 1420..1503  
 FT /note="acyl carrier protein domain, ACP6"  
 Domain  
 EP-791655-A2.  
 PD 27-AUG-1997.  
 PD 19-FEB-1997; 301056.  
 PR 22-FEB-1996; US-012078.  
 PR (ELIL) LILLY & CO ELI.  
 PI Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL; WPI: 97-418046/39.  
 DR N-PSDB; T80414.  
 PT DNA encoding Streptomycetes fradiae tyactone synthase domain - for production of tylosin-related polyketide compounds  
 PS Example 2; Pages 167-172; 220pp; English.  
 CC W22606-W22610 represent proteins encoded by the platenolide synthase gene cluster. The gene cluster is also referred to as the smg gene, and was isolated from Streptomycetes ambofaciens. These sequences are multi-functional proteins which direct the synthesis of the polyketide platenolide. Platenolide is the basic building block of the macrolide antibiotic spiramycin. The DNA encoding this sequence was used along with the tylg gene (see T80413) to create a hybrid ORF1 sequence (see T80415). The tylg gene is the tyactone synthase gene cluster of the invention.  
 CC The tylg sequence was isolated from Streptomycetes fradiae, and encodes multifunctional proteins which direct the synthesis of the polyketide tyactone. Tyactone is the basic building block of the antibiotic tylosin. The hybrid sequence can be used to transform S. ambofaciens lacking the smg ORF1 sequence, or S. fradiae lacking the tylg ORF1 sequence, so that they can produce polyketides. The DNA sequence can be modified so as to alter the type of carboxylic acids incorporated, the number of carboxylic acids incorporated and/or the post-condensation reactions performed, thereby resulting in novel tylosin-related polyketides.  
 SC Sequence 1580 AA;

Query Match 87.5%; Score 42; DB 26; Length 1580;  
 Best Local Similarity 85.7%; Pred. No. 1.11e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1255 ggvlgk 1261  
 QY 1 GTVLGK 7



QY 1384 SKTSKV-V-YDVAVDDTTIEVKDKKLGVTTLTSTGTG-ANKFALSNQATGDALVKASD 1440  
 DB 500 NALD-NSLEFDSGAKTSTVATSGSKTLVKGAEVETLVNIDTAFNGADVSFGKDAQ 558  
 QY 1441 IVAHLNLSDDICTAKASQASNSAG-YVDADNKV-TIDSTNKKYQAKNDGTVTKNE 1498  
 DB 559 SGKFSVKTGDKIEFVGTTLEGSVIDAGNDTIAMKSAALTSANFTMIKNIENVAIS 618  
 QY 1499 VAK-D-KLVAAQATPDGTLAQNMVKSVINKEQVNDANK-QGINEON-AFVKLEKASD 1554  
 DB 619 DAVATADLSSAKRNTITTKKE-AADTTLINK-DOYINTAADRSVKILTKLDVYG 676  
 QY 1555 NKTNAAVTGDINAAVQATPLTFAGDTGTAKKLGELTLTKGGQDTNKTLDNNIGVAVG 1614  
 DB 677 AND-VAVTIVDAAKDASIALGTEATDKALV-IDTGLETINTSLVATSPENTANTV-N 733  
 QY 1615 TDGTTVALADLTNLSVNVNAGTKIDDKGVSPVDSGQAKANTPVLSANGDLGKVISN 1674  
 DB 734 -AKLT-DVTSIID--GDAR--ITLGHAGTAGTDYSKVSMIDAR-ALKAGLTFDASATL 786  
 QY 1675 VGKGTOKTDANVQQLNEVERNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSSNRTVTKA 1734  
 DB 787 GANATTGGSGADSITVKGNIYVDL--VAGD-DITLKGKREKDDITVNNFNAGDKI 843  
 QY 1735 GTVLGGKNDTEKLAGVQVGVKDGANGLSNVWTKQDGSKALLATYMAAGOT 1794  
 DB 844 D 844  
 QY 1795 N 1795

RESULT 15  
 ID 048031 PRELIMINARY; PRT; 1536 AA.  
 AC 048031;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)  
 DE ADHESIN.  
 GN HMW1A.  
 OS HAEMOPHILUS INFLUENZAE.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 OC PASTEREILLACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-12;  
 RX MEDLINE: 92192797.  
 RA BARENKAMP S.J., LEININGER E.;  
 RL INECT. IMMUN. 60:1302-1313(1992).  
 DR EMBL: 00876; G473771;..  
 SQ SEQUENCE 1536 AA; 159916 MW; 140BBE9B CRC32;

Query Match 1.6%; Score 217; DB 2; Length 1536;

Best Local Similarity 18.7%; Pred. No. 1,08e-10;  
 Matches 154; Conservative 242; Mismatches 372; Indels 54; Gaps 51;

DB 668 LNVSEGEFNTLIDSRGSDAGTLQPYNLNGISFNKDTFNVERNARVNDIKA-PIGI 726  
 QY 1282 LTVGNNGKGIIVDSONGQNTIT-GLSNLTANVTNDKGSVRTTEOGKIIXDEKTRASI 1340  
 DB 727 NKYSLLANA-SFNGN-IS-VSGGGSVDFT-LLASSNVQPGVINSKY-FNVSTGSSLR 781  
 QY 1341 VDVLASGNLQNGEAVFVSTYDVFNFADGNATTAKVTYDITSTKSVYDVAVDDT-T 1399  
 DB 782 FKTSGSTGFSIEKDLTLNAGNITLLQVEGT-DGMI-CKGIYAKKNITFEKG-NITF 838  
 QY 1400 IEVKD-KKLGVKT-TTLSTGTGANKFALSNQATGDALVKASDIYAHLN-TLSGDIQTAK 1456  
 DB 839 GSKRAVTEIEGNTVNNANVTLLISDFDNHQ-KP-LTIKKDVIILNSGNTLAGNIVNIA 896  
 QY 1457 GASQANSS-AGYVDADGN-KVIYDSTDKYQAKNDGTVDKTEVAKDKLVAQATPDGT 1514  
 DB 897 -GNLVEANANFKATNTFNVGGLFDNKG-NSNISIAKGARFKDIDNSK-NLSTTNS 953

QY 1515 LAQNVAVSVIN-KEQVVDANKKQGINEDNAFPVKGLEKASDNKTKNAAYVGLNMAVQ 1573  
 DB 954 SSTYKTIISGNITKNKD-LNITNEGSDTEMOJGQVSOKEGNLTISSDKINITKQITIK 1012  
 QY 1574 PLTF-AGDTGTAKKLGELTLTKGGQDTNKTLDNNIGVAVG--TDGF-TVKLAKDLTNL 1629  
 DB 1013 AGVDGNSDSDATNANLITIKELKL-TODLNISGN-KAEITAKGSLITIGNINSAD 1070  
 QY 1630 NSVNVGKTKID-KGVSPVSSGQAKANTPVLSANGDLGKVISNVGKGTOKTDANVQ 1688  
 DB 1071 GTNAKVTENOVKDSKISADGKVTLSKVTSGSNNTEDSDNNAGLTIDAKNTVNN 1130  
 QY 1689 QINEVRNLLIGCNAGNDNADGNQVNT-ADIKKDPNSGSSSNRTVTKAGTVLGGKNDTE 1747  
 DB 1131 NITSHKAVSISATSGELTTKTGTINATGTGNETTAQTSILGIESSGSVTLTATEGA 1190  
 QY 1748 KLAT-GGVQGVKDGANAGDLNVMWVKTK-D-GSKK-ALLATYNAAGOTNLTNNPAE 1803  
 DB 1191 LAVSISGNTYV-TANGALTLTLAGSTIKGESVTSSQSDIG--GTISG-GTVEV-K 1245  
 QY 1804 -AIDRINQGIREFHVNDNGQEPVQGR-NGIDSSAGKHSVALGFQAKADGEAVALGR 1861  
 DB 1246 ATESLTTQSNK-IKA-TTGEANTYSATGT-GGLISGNTVY-VTANAGD-LTVNGAEI 1300  
 QY 1862 QTA-GNOSIALGDNAGATGOSTALIGTVNACKHSGALGDPSTVAKADNSYSVGNNOF 1920  
 DB 1301 NATEGAATLTTSSGKLTTEASSHITSAGOVNLSAQDSVAGSVAANAVNLNTTGLTIV 1360  
 QY 1921 PDATQDVGVGNNTVTVESNSVALGNSAISAGTHAGTQAKKSDGTAGTTTAGATGV 1980  
 DB 1361 KGSNINATSGTLVNAKAE--LNGAALGNHTVYVNTANGSGSVIATTSRVNITGLI 1418  
 QY 1981 KGFAGQTAAGAVSAGSAERLQNVAAAGEVSATSTDAVNS-OLYKATQGINATNELD 2039  
 DB 1419 -TINGLNIISKNGINTVLLKGVKIDVKYIOPGIASVDEVIEA 1459  
 QY 2040 HRHONENKANNGISSAAMAMPQA-YI-PGRSMVTGIAAT 2079

Search completed: Wed Oct 28 11:21:06 1998  
 Job time : 471 secs.





```

Db 523 FGDTSADGSGOLSVKTAGNDTLNLEAKL--KAGSLIDGEGNDITLIMKASALADAATL 580
QY 1380 Y-D-DTSEKTSKYVDVAVDDTTIEVKDKKLGVTTLTSLTGANKFALSNOATGA--L 1435
Db 581 GMIKNI-ENV-TVSDALSANTDVASASSEVNIIGLL-ADKTAPEPELYNNK-QIIDDOSKE 636
QY 1436 VKASDIVAHLTLTSGDIQTAKGASQAN--SSAGYVDADGNKYIYDSTDNKKYQAKNGTVD 1494
Db 637 MAKSOILLIKMNDMSGDDIYVNIYLNKATINOGDKRVAAAGQTKGKIDGIESVNIIS 696
QY 1495 KTK-EVAKDKLVAAQTPDGTGLAOMNKVSVLNK-EQ-VDNANKKQGINEDNAVYK-GLER 1550
Db 697 VAKDNTTANTLIMIDTSSDGTGKANKIYISGDDITVAASVATGKSIKDLASALTGKL 756
QY 1551 AASNKTKRNA--V-TVGD-LAAVAQTPLTFAGDTGTAKKLGTL-TIKG-GQTD-TNKL 1604
Db 757 TFDASVANKLAS 767
QY 1605 T-DNNIGVAVG 1614

RESULT 11
Q48028 PRELIMINARY: PRT: 1477 AA.
Q48028; 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
GN ADHESIN.
OS HAMOPHILUS INFLUENZAE.
OC PROKARYOTA: GRACILICUTES; SCOTOBACTERIA: FACULTATIVELY ANAEROBIC RODS;
OC PASTEURILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-12:
RX MEDLINE: 92192797.
RA BARENKAMP S.J., LEININGER E.;
RL INFECT. IMMUN. 60:1302-1313(1992).
DR EMBL: U08875; G482841;
SQ SEQUENCE 1477 AA; 154472 MW; FEA8950 CRC32;

Query Match 1.8%; Score 244; DB 2; Length 1477;
Best Local Similarity 19.2%; Pred. No. 4,92e-14;
Matches 154; Conservative 229; Mismatches 373; Indels 48; Gaps 46;
Db 584 DANKAKIVAGTIT-TGEGKDFRANVSL-NGTGKGLNISSVNNLTHLSGTINISGN 641
QY 1255 ENGITTAKVKKVYRGIDOTKGLTPKLVGNNGKGI-VIDSONG-QNTITGLSNTLAN 1312
Db 642 ITINOTTRKNTSIWOTSHDSHMNVSALEET-GANFTFKITISSNSKGLTTOYRSSAGVN 700
QY 1313 VTNDKGSVTRTEGKLIKEDKTRAAISYDVLSAGFN-LQ-ONGEAYVFVSTYDTVNEAD 1370
Db 701 FENGVNMGSEFN-LKEGAKYFKLKPNEN-MNT-SKPLPFRPLANTATGSGSVFEDIYAN 757
QY 1371 GNTTAKVYVDDTSKTSKYVIVNDVDTIEYKDKLGVK-TTTLTSLTGANKFAL-SN 1428
Db 758 HSGRGAELKMSSEINISNGANFTLNHVR-GDDAFKINKDLT-INATNSNFSLRQTKDFY 815
QY 1429 QARGDALVVASDI-VAH-LN-TLSGDIQTAKGASQANSAGVYADGNKYIYDSTDNKKY 1485
Db 816 DGYARNAISYVIST--LGGNV-TLGGQSSSITGNTTIKAAVITLLEANNANNOONI 872
QY 1486 QAKNDGTIVKTEKAVDKLVAAQTPDGTGLAOMNKVSVLNKQVNDANKKQGINEDNAFV 1545
Db 873 RDEVIKIGSLVNGSLST-GE-NADIKGNLTIS-ESATFKGTRPDLTINFTNNGT- 928
QY 1546 KG-LEKASPNKTKNAAYVYGDILNAVAQTPPLFAGDGTGTAKKLGTLTIKGGQDNTKL 1604
Db 929 AEINITGVVKG-NTYNDGL-NITT-HAKRNORSIIGGDIINKKSGLNITDSNDAD- 984
QY 1605 TDNNIGVAVGTFVKLAKDLTINLSVNVAGGTIKDKGVSFVSSGQAKANTPYLSANG 1664

```

```

Db 965 IOIGENISQKSGNLTISDKINITKQITIKKIDGEDSSSDATSNANLTJTKELKLED 1044
QY 1665 IDLGKGVISNKGKTKEDDANV--QOLNEVNNLLGLNAGNDNDGN-QVNIADIK-KDP 1721
Db 1045 LSISEFNKAETAKGREDLTIGNSNDGSGAEAKTVTFENNKKSIADGHNTVLSKRV 1104
QY 1722 NSGSSSNRT-YI-KAGTYLG-GKGNDEKELATGGVGVYDCKDNANGDLSNVWVKIQ-K 1777
Db 1105 TSSNGGRES-NSDNDTS-LITT-AKNE-VNKDITSLKYVNIITASEKYTTTLAGSTINAT 1160
QY 1778 DSKKALLATYNAAGQVYLTNNPAAIDRIEDGIRFFHYNDGNOBPVVOGRNGIDSSA 1837
Db 1161 NKASATTK-TGDISGITSGNTVSVSATGDLTATSGSKIEAKSGEANTSATGTI-GGTI 1218
QY 1838 SGKHSVAIGFQKADGEAVALIGROTONGOSIALGDNAQA-TEDQSIALGTGVNAGKH 1896
Db 1219 SCNTYV-VTANAGD-LTVNGCAEINATEGAATLPAATGTLTTEAGSSITSTKGVDLLAQ 1276
QY 1897 SGALGDPSTYKADNSYVSGNNQFTDATQTFVGGNNITVTSNSVALGSNSAISAGTH 1956
Db 1277 NSGIASINAAVTLNTTGTLTVAAGSDIKATSGTLVYNADA--KLNGDASGDSFEVNA 1334
QY 1957 AGTOAKKSDGTAGTTTGTATGTVKGFAGQTAAGVAVSGASGAERRIQNVAAGEVSATST 2016
Db 1335 VNASGSGVTAATSSSVNITGDLN 1358
QY 2017 DAVNGS-QLYKATQGIANKATNELD 2039

```

```

RESULT 12
ID 052647 PRELIMINARY: PRT: 834 AA.
AC 052647;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE PROTEIN A (FRAGMENT).
GN OMPA.
OS RICKETTSIA AKARI.
OC PROKARYOTA: BACTERIA: GRACILICUTES; SCOTOBACTERIA:
OC RICKETTSIA AND CHLAMYDIA; RICKETTSIALES; RICKETTSIACEAE;
OC RICKETTSIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KAPLAN:
RX MEDLINE: 93194085.
RA GILMORE R.D., JR.;
RL GENE 125:97-102(1993).
DR EMBL: L01461; G152484;
DE OUTER MEMBRANE; REPEAT.
FT NON_TER 1
FT NON_TER 834
SQ SEQUENCE 834 AA; 79349 MW; 21476055 CRC32;

Query Match 1.7%; Score 225; DB 2; Length 834;
Best Local Similarity 19.9%; Pred. No. 1.13e-11;
Matches 158; Conservative 188; Mismatches 400; Indels 48; Gaps 45;
Db 25 KATITDITDAASVLTITLNAVAVLTGVYDNT-TGVNVGVVLNNGALSOYTGNGN-TNSL 82
QY 718 KTTADDTLTKNNIGVAVDSTNSLTIVKAKTSLDCLDVNTKTLTAS-DKTYVDSGNNTATL 776
Db 83 ATISVAGATLIGAVIKATTTDLTDAASVLTITLNAVAVLTGAIDNTGVGVNVGLNNG 142
QY 777 QNGDLTFSKONTGATPATNTKT--IYVGDG-LKFTDNNIGALDGTIYTK-DKVGFAKODG 833
Db 143 ALSQVTGNIGNTN-SLATISVAGATLIGAVIKATTTDLTDAASVLTITLNAVAVLTGAV 201
QY 834 SLDSKPYLDKDKLAKVEVEITTINGINAGKATGTGSLNTLTDAFTNATGTVGLGIVDT 893
Db 202 DNTGVNVGVVLNNGALSOYTGNGNLSLATISVAGATLIGAVIKATTTDLTDAAS 261
QY 894 DKTRASIGVVLNAGFNLKNNNGAKDKEVSTIYDIYDFINGNATIAKTYTGDKASKAVAYDVN 953

```







Db 193 NINGGIIITVDAINGTIGNTALATVNVGAGIATLEGAIKATTTKLTNASVLTLT 252  
 QY 580 NTNSKQIYOVGADGITFIDISNKPAGIEMTRITR-DGIGFANNGSLDANKRPLRPT 638  
 Db 253 NVNAV---LTGAIDTGVNDVGVNLNGALSQVTGIGNTALATISVSGAKATLGAV 309  
 QY 639 GINAGKEITVNSAIPNATNGQJLDFMNRLS--TANTEKSGSATITKIDLYNLQVPLTE 696  
 Db 310 IKATTTKLTDNASAVETNPVYVGTGIDTGNANGIYFTGDSVTGIGNT-NALATV 368  
 QY 697 AGDTGPNTKIGEL-LKVGSKTTR-DDLTKNNGIYVADSDTSLYKALKLISDDAV 754  
 Db 369 NVGAGELLOYGVKANTINLTDNASAVETNPVY-VTGALDN-TGNANGIYFTGDS 426  
 QY 755 NTKT-LTASDKVTVDGNTAKLQNGDLTFESKONTGATPATNSKTIYGVDTL-KETDNG 812  
 Db 427 VTGNIGNTALATISVSGAKATLGAIKATTTKLTDNASAVETNPVYVGTGIDTGN 486  
 QY 813 IALD-GTIV-IRKDVYEFKQ--DGLSKRPY-L-DKDK-LKGEVETTTGILNAGSKA 865  
 Db 487 NNGIYFTGDSVTGIGNTALATISVSGAKATLGAIKATTTKLTDNASAVETNP 545  
 QY 866 ITGLSTLTLDAL-NATGIVTOLGIYDSDTKRRASIGVYLNAGF-NLKNNGDADFVST 923  
 Db 546 VVVTGALDNTGNANGIYFTGDSVTGIGNTALATISVSGAKATLGAIKATTT-K 604  
 QY 924 YDVTDFI-N-GNATTAKYVDGKASKVADYVVDG-TTILHGTAGDNKNQIGVKTTLTK 980  
 Db 605 LUDNAS-AVETNPVYVGTGIDTGNANGIYFTGDSVTGIGNT-NALATV-N-GA 660  
 QY 981 TPAKGGKALNEFVNSGDDKALINAKIADNLTLAGIINTGTADTALQTVQVKKVEN 1040  
 Db 661 GIATLEGAIKATTTKLTDNASAVETNPVYVGTGIDTGNANGIYFTGDSVTGIGNT 715  
 QY 1041 GDDNDADITVYKDAKTNOVNTLKLKGNKGLDIOFNKDGITVFGINTQSGKAGNNTL 1100  
 Db 716 SQVTGIGNTALATISVSGAKATLGAIKATTTKLTDNASAVETNP-PVYV-TGALDN 773  
 QY 1101 NNGSLIKNTAGNEQIYOVGADGVKFA-KVNNGVGAGIDGTIRITDEIGEPATGSLDK 1159  
 Db 774 T-GN-ANNGIATFTGDS-T-V-TGDIG-NTN-ALATISVSGAKATLGAIKATTT-K-L 824  
 QY 1160 SKHLSKDGINAGGKITTINQSEIANSNDATVGGKIYDTKLELNKISSRAKTAQNL 1219  
 Db 825 TDNASAVETNPVYVGTGIDTGNANGIYFTGDSVTGIGNTALATISVSGAKAT 883  
 QY 1220 HESVADDEQNNFTVSNPSSYDTSKIDTIPAGENGITTKVKNKGVVRGIDOTKGLTT 1279  
 Db 884 LGGAIKATTTKLTDNASAVETNPVYVGTGIDTGNANGIYFTGDSVTGIGNT 943  
 QY 1280 PKLTGVNNGKGIYIDSQONNTITGL-SMTLANVTNDKGSVPT-TEQGTIKIDEDKTRA 1337  
 Db 944 LATVNV-GAGVTLQAGSLDANNIDFAGR-STLEFNGPLDGGGKAIPYEFKALANGNNA 1001  
 QY 1338 ASYDVLSAGFNLOGNS-BA--VDVSTYDYVNFADGNATTKA-VTYDOSTKTSVYVD 1392  
 Db 1002 I-LNVNTKSLASHLTIGTVAELINIGAG-NLEAIDASAGDVTLTLMODIHFRALDSALV 1059  
 QY 1393 VNVDDTIEVKDKLGVKTTTLTSTGANKFALSNOQTALDALKASDIYAH-LNT-LEG 1450  
 Db 1060 SNTLGVVNNILLAADLYAGVDEGVYFPGVGVNLTGNSVNAARNIDVCGNFENL 1119  
 QY 1451 DIOTAKAQSANSAGYVDA--DENKVIYDSTDKKIYQAN-DEYVDKTEVAKDKLYAQ 1507  
 Db 1120 LIYNAVITTD-DVN-LEGIQNVL-INNNADFTSSTAFNACTIOINDATYITDAN-N-GNL 1174  
 QY 1508 AQDTGDTLQANVSVYKNEQVNDANKKQGINEDNAFVKLEKASPNKTKNAAYVYVGD 1567  
 Db 1175 N-IPAGNIORHADAQILLONSSGNDRTITLGANIDPDN-DDGVIYLNSTVTKGLTIA 1232  
 QY 1568 NAVAQDTLTPFA-GDTGTT-AKKIGETLTK-GGQTDITNKLJTDNNIGVAG-TDGFTVYKLA 1623  
 Db 1233 GGTFTGGAHKLQDITVFKGEG-DFGTAGTTFNTNIVLITGQLELGAAT-ANVVL-FKD- 1288

QY 1624 KDLTNLSVNAAGCTKIDKGVSEFVDSGQAKANTVYLANG-LDLAGVANSVGCTGDT 1682  
 Db 1289 -AVOLTQTNIGIGDFENAKNGTIVLNNVNAVGTAKNT-GGTNGGTLIVGASMLN-RV 1345  
 QY 1683 DAAVQOLNEVRNLLGLGNAGNDNDGNOVNTADIKDPNSGSSNRPYIKAGTVLGKRG 1742  
 Db 1346 NG-IAMKVGAGNVTYAKGKNKIGFIOGTGNTLTPRAHFELTOSINKTGGQALKLFM 1404  
 QY 1743 NNDTEKLTAGGVQVGVDKGNAN-GDLSNWWKTQKDGSKKALATYN-AGQTYVLTNN 1800  
 Db 1405 NGSVSGVGTAAAN--SVGDITTAGATSPASSNAKGAATLIGTTSFANTFTNTGAVTLA 1462  
 QY 1801 PAAIDRINEQIRFRHVNDGNEPVOGRNGIDSASGKSHVALGFOKADGEAAVAG 1860  
 Db 1463 KGSITSE-AKNYATSEFVANSATINFGSLAFNSNTGS-GTTLTGANQVYTT-GTGS- 1518  
 QY 1861 ROTQAGNOSIAIGDNQATGDSIAIGTGNVYAGKHSAGIDPSTVKADN-SYSVGNNO 1919  
 Db 1519 FTDTLNTLTFDGAASGGNIIKSGSTLIDGVSMLALV-TATNFDMNNTSPKRYTV 1577  
 QY 1920 FTDA-TQTDVF-GV--GNNTVYESNVALGNSAISAGTAGTQAKKSDGTAGTTTA 1974  
 Db 1578 ISAFAGLKL-PTPKENYKITTINDNR 1603  
 QY 1975 GATCTVAGFAGOTAVGAVSVGASGAER 2001  
 RESULT 5  
 ID 054356 PRELIMINARY; PRT; 831 AA.  
 AC 054356;  
 DT 01-JUN-1998 (TREMBLREL, 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL, 06, LAST SEQUENCE UPDATE)  
 DT 01-DEC-1998 (TREMBLREL, 07, LAST ANNOTATION UPDATE)  
 DE HIGH MOLECULAR WEIGHT OUTER MEMBRANE PROTEIN.  
 OS USPAL.  
 GN MORAXELLA CATARRHALIS.  
 OC PROKARYOTA: GRACILICUTES: SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
 CC NEISSERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-035E;  
 RX MEDLINE; 98013056.  
 RA ABEI C., MACIVER J., LATIMER J.L., COPE L.D., STEVENS M.K.,  
 RA THOMAS S.E., MCCracken G.H. JR., HANSEN E.J.,  
 RL INFECT. IMMUN. 65:4367-4377(1997).  
 DR EMBL; U57551; G2772586;  
 SQ SEQUENCE 831 AA; 88278 MW; 6527B5CA CRC32;  
 Query Match 2.1%; Score 281; DB 2; Length 831;  
 Best Local Similarity 34.9%; Pred. No. 9,00e-19;  
 Matches 75; Conservative 54; Mismatches 69; Indels 17; Gaps 15;  
 Db 122 GSNNETATNEYSTIVGDDNKATGRYSTIGGDNNTREGEYSTVAGKKNQATGTGSPAG 181  
 QY 1832 GIDSSASGKSHVALGFOA-KADGE-AAVAIG-ROTOAGN-OSIAIGDMQATGDSIAIG 1887  
 Db 182 VENQANMANAVAKKNIIEGENSVIAIGSENT-VKTEHKVNFILSGTGTGVSNSVLLG- 239  
 QY 1888 TGNVYAGHSGAIDPSTVSKADNSYSVGNNOPTDAOTDVFYGNNTIYTESNSVALGS 1947  
 Db 240 N-E-TAGQATY-VKNAE-VGLSLT-GFAGESEK-A-ENGV--SVVSGEGGERQIVNG 289  
 QY 1948 NSAIAGTHAGTQAKKSDGTGTTTATGATGTGKGFAGTAVGAVSVGASGAERRIONVA 2007  
 Db 290 AGOISDSTDAVNSQALHALTVYDDNDQYDVLVNR 324  
 QY 2008 AGEVSATSTDAVNSQALYKATOGIATATNEL-DHR 2041  
 RESULT 6  
 ID 022248 PRELIMINARY; PRT; 1845 AA.  
 AC 022248;

```
QY 1755 QYGV-D-KDGN-NGDLSNVWKTQKDGSKKALLATYNAGCTNYLTNNPA--E--AID-R 1807
Db 759 YODGKGVSVYDNTGATITTKNGSGYVTGNQVADALAKSGFELGLADEADAKRADDKTK 818
QY 1808 INEÖGRFHFVADGNOEPVQGRNGIDSSASGKSHVAI-GFO-AKAD-CEAVALIG-R-Q 1862
Db 819 ALSAGTELEVNAHDKYFRFANGLTIVYSAATVSTIDANGKATYTTVKVDLPLQIYNT 878
QY 1863 T-QAGNOSTAIG-DNAÖATGDSIAIGTGNVAKHSGAIGDPSTVKAD-N--S--YSV 1914
Db 879 DANGKIKTKVVKDQGTQWYELNADGTADMTKEVLGNVDSDKKVVYKNDGKWKYAKADG 938
QY 1915 -CNNNOFTDAT-QTDV--FGVGNITVTESNGVALGSNSAISAGHAGTQAKKSDGTA-G 1969
Db 939 TZDKTKGEVSNKYSSTDEKHAVYSLDPNDOSKKGKVVYIDNVANGDISATSDAINGSQLYA 998
QY 1970 TT-TTAGATGTGKFPAGQTAAGAVASVGA-S-GAERRIONVAAGEVSATSTDAVNSQLYK 2026
Db 999 VAKGYTNLAGVYNNLEKGNKVKRADAGTAGALASOLPOATMFGKSMVMAIAGSSYQO 1058
QY 2027 ATÖGIAN-A--TNELDHRHÖNENKANAANGISSAMAMASNPÖATYIPGRSMVTGATHTNGQ 2083
Db 1059 NGLAIGVSRISDNGKVIIRLGGTNSÖQKTYAAGVGYO 1097
QY 2084 GAVAVGLSKLSDNGQWVFKNKSGADTQGHVGAAGAFH 2122

RESULT 3
ID 052708 PRELIMINARY; PRT; 1029 AA.
AC 052708;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DE 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE PROTEIN A (FRAGMENT).
GN OMA.
OS RICKETTSIA CONORII.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; RICKETTSIAS; RICKETTSIALES;
OC RICKETTSICEAE.
RN [1]
RC SEQUENCE FROM N.A.
RK STRAIN-KENYA TICK TYPHUS;
RX MEDLINE; 93194085.
GX GILMORE R.D. JR.;
GB GENE 125:97-102(1993).
EMBL: L01462; G152486; -
KW OUTER MEMBRANE; REPEAT.
FT NON_TER 1
FT NON_TER 1029
SQ SEQUENCE 1029 AA; 99999 MW; A036CADD CRC32;

Query Match 2.2%; Score 291; DB 2; Length 1029;
Best Local Similarity 20.4%; Pred. No. 4.44e-20;
Matches 191; Conservative 244; Mismatches 448; Indels 51; Gaps 48;

Db 49 GAIIDNTT-GVDNVG--VLINIGALSQYT-GNIGNTNALATI-SVGAAGKATLGAVIKAT 103
QY 850 GEVEITTINGINAGKATITGSLNTLIDATNFTGHTVQLGIVDSTDKTRASIDVILNAGF 909
Db 104 TKLTDNASQYTFNPNVYVYTGAIIDTGNANGVIYTFGDSVTYGNIGNTNALATISVGAAG 163
QY 910 -NKKNNNDARDFSTYDVTDFI-N-GNATTAKYTDGKASKAVAYDVVDG-TTIIHLTGAD 965
Db 164 ATLGGAIIKAT-TKLTDNASQ--VTFNPNVYVYTGAIIDTGNANGVIYTFGDSVTYGNIG 221
QY 966 GKNKÖIGVKTITLTKDADKADKAINFSVSGDKALINADIDNNTLAGETIRNKGTA 1025
Db 222 NTNALATISVGAAGKATLGAIKATTTKLTDNASQYTFNPNVYVYTGAIID-NTGNANGVIY 280
QY 1026 DT-ALÖFQYKVKYKEN-GDDNDADADITVYGKA-KTNÖVYVTLKKKNGIDIOITNKDGY 1082
Db 281 TTFGDSVTYGNIGNTNALATISVGAAGKATLGAIKATTTKLTDNASQYTFNPNVYVYTG 339
QY 1083 TF-GINTÖSLGKAGNNTLNNGLST-KNTAGNEÖIÖVGADGVK-PA-KVN-NGVYVAG 1136
```

```
Db 340 AIDNTGNANGVIYTFGDSVTYGNIGNTNALATISVGAAGKAT-LGGAIIKATTTKLT 398
QY 1137 -IDÖTFTITDEIGFACIÖNSLÖKSRPHLSKÖDINAGKKTITÖSGET-AÖNSNDAYTG 1194
Db 399 SQYTFNPNVYVYTGAIIDTGN-ANGVIYTFGDSVTYGNIGNTNALATVYVAGGLRVÖGG 457
QY 1195 GKI-YDKTLELKNISSTAKTÖNSLHEFSVADÖQNNFTVSNPYSYDTSKTSVDITFA 1253
Db 458 VYKSTINILDNASQYTFNPNVYVYTGAIIDTGNAN-NGIYV-F-TDSDVTYGNIGNTNL 514
QY 1254 GENGITTKVKKVYRAGIDÖTKÖLTPKLTGKNNKGIYIDÖSÖNÖMTIG-LSNTL 1312
Db 515 ATISVGAAGKATLGAIKATTTKLTDNASQYTFNPNVYVYTGAIIDTGN-ANGVIYTFG 573
QY 1313 VTNDKGSVFTTEÖGKIIKÖDÖKTRASIDYV-LSAÖFNÖQNGEANDPSTYDITVNFADG 1371
Db 574 STYTGNTGNALATVYVAGGLRVÖGGVYKSTINILDNASQYTFNPNVYVYTGAIIDTGN 633
QY 1372 NATTAKYTDÖDTSKTSKYVDV-NVDÖTIEYKDKKLGKTTTLSTGÖANKFALSÖQ 1430
Db 634 NANGGI-VT-FTGD-STYTGNTGNALATVYVAGGLRVÖGG-VY-KSNTINILDNASQ 688
QY 1431 TÖGALYKASDIYAHNTLSGDIÖTAKGASÖASSAGYVDADNKKYIDSTÖDKRYÖARKND 1490
Db 689 VTFNPN-VYVYTGAIIDTGNANGVIYTFGDSVTYGNIGNTNALATVYVAGIATLEGAVI 747
QY 1491 GTVDKTEVAKDKLVÖAQÖTPÖTILÖQNVKSVINKEÖYNDANKKÖGINEDNAPYK-G-LE 1549
Db 748 KATTTKLTNAAVYLTNNVAVLTGAIIDNTTGVNVGVLNIGALSQYTGNTGNALAT 807
QY 1550 KAASDKTKRNAAV-TVGDNLNVAQÖPL-TFAGDGTITAKKLGÖETIL-TKGÖDÖTNRKLT 1606
Db 808 ISVGAAGKATLGAVIKAT-TKLTDNASQYTFNPNVYVYTGAIIDTGNANGVIATFGDS 865
QY 1607 NNIGVYAGDGTFTYKAKLTILNVSYNAGTÖKIDKÖVS-EYDSSÖKAKANTPVLSANGL 1665
Db 866 TYTGNTGNALATVYVAGGLRVÖGGVYKSTINILDNASQYTFNPNVYVYTGAIIDTGN 925
QY 1666 DLGKVIYVNGKGTÖDÖTDAVYÖ-ÖLNEVYR-NLGLGNAÖNÖMNDGÖNVNADIKKÖDPS 1723
Db 926 ANNGIYTFGDSVTYGNIGNTNA--LATISVAG 957
QY 1724 GSSNRTYIKAGTYDLGKGNÖDTERLATGVQVÖG 1757

RESULT 4
ID 052657 PRELIMINARY; PRT; 2021 AA.
AC 052657;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DE 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE 190-KDA ANTIGEN (ROMPA).
OS RICKETTSIA CONORII.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; RICKETTSIAS; RICKETTSIALES;
OC RICKETTSICEAE.
RN [1]
RC SEQUENCE FROM N.A.
RK STRAIN-MALISH 7;
RX MEDLINE; 94171067.
GX CROCÖUET-VALDES P.A., WEISS K., WALKER D.H.;
GB GENE 140:115-119(1994).
EMBL: U01028; G467814; -
KW
FT
FT
SQ SEQUENCE 2021 AA; 203366 MW; 20676FF2 CRC32;

Query Match 2.2%; Score 299; DB 2; Length 2021;
Best Local Similarity 20.0%; Pred. No. 3.93e-21;
Matches 306; Conservative 391; Mismatches 726; Indels 104; Gaps 93;

Db 136 INAGKSLTLNÖNNDAAHGEFADAPADYVYTG-LGNIALG-GANALITÖSAPAKITLAG- 192
QY 520 ITVGNKNNPDÖQYVNTLTKIKENGÖVÖDVTTFETNGIYVTEGLNÖNNGIYVGNSTLNNDGSLVK 579
```



QY 765 TVPDSGNNTALQNG-DLFFSKONTGATPATNSKTIYGVDDKLTFTDNNKSLDSTYITK 823  
Db 387 KKIY-A-DTTLALTYGGVVA-EIAK-ED-D-KKLVNAGD-LVYALON-LSMKAKABADT 438  
QY 824 DKVGRKDDGSLDKSKPYLDKDKLVGEVEITTINGINAGKATIGLSLTLDATNATG 883  
Db 439 DGALEGISKDEVKAGETV-TEKAGNKLKVKODGANF-TYSLDALTGLSLTIGSTTN 495  
QY 884 VTOLGIYSTDKTRASIGDVLMAGFNKKNNGDAKDPSTIDYDFING-NATT-ARVY 941  
Db 496 GGNDAKIV--INKDGLT--TPA-GNGGTGTNTISVTKDGIKANKAIT-NVASGL-RA 548  
QY 942 DGKASKVAYDVNDGTTHLTGADGNKNOIGVTKTTLTKTDK-GDKAINTSVNSGDKA 1000  
Db 549 YDDA-NF-DVLNKSATDLNRHVEDAKKLLNLENNAKOPLV-TDSTAAVVGSLRKLGW 605  
QY 1001 LIMAKDIADNLTLAGELRNTKGTADTALQTFYQKRVKENGDDDDADTLTVGDAKTNO 1060  
Db 606 VVSTK-NGTKESNNGVKADEVLFTGAGAAVTSKSENGKHTIVVAETKADGLEKDG 664  
QY 1061 VNTLKLKKNNGDIOTNNKDGVTTFEINTOSGLKAGNNTLNNGLSIKNTAGNDQIOGA 1120  
Db 665 DTIKL-KVDNONTDNLTVGNNNGTAIVTKGFEYTKGTADRGKVYKADATADAKV 723  
QY 1121 DGKFAKAVNGVAGADID-GLTRITRDEIGFAGTN-GSLDKSKPHLS-KDGI-NAGGKRI 1176  
Db 724 AIVK--DVATATNSAATPVKTEKNTLSTIDED-NPTDNCKDALKAGDITLFRKAGNKLK 780  
QY 1177 TTIQSGELIANSNDAYTGKITYDLKTELENKISTAKTAONSLEHF-SVADGONNFTVS 1235  
Db 781 RDGKNTEDLAKNLEKVT-AKVSIDL-T-IGNTPGTGTAAPKVNITSTADGLNFKETA 838  
QY 1236 -NP-YSSVSTKSTSDVITPAGENGITTKVKNKGVAVGIDQTKLTPLKLVGNNNGKGI 1293  
Db 839 -DASGKNVYLGATITLLE-PSA-GA-KSHVDLVNDAKTSMAASIEDVILRAGMNNG 894  
QY 1294 IDSONOQNT-ITGLSNTLANTVNDKSVRTTEOGKITIDEDKTRASIVDVLSAFENLQ 1352  
Db 895 NGNNDVYATVPTVNEFTDSTGTTVTYQKAD-GKGA-DVKIGAKTSVYKDHNGKLTG 952  
QY 1353 NGEAVDFSTYDTVNFADONATTAKTYTDDTSKGVYDVAVDDTLTEVKKKGY-KT 1411  
Db 953 KDLKANNAGTVSEDDGKDTGLVTAKTVIDAVNK-SGWRVYTGEGAT-AETGATVANG 1010  
QY 1412 TLTISTGANFALSNOATGADALVAKSDIVAHMLTSLGDIOTAKGASQANSAGYVAD 1471  
Db 1011 NAEVTVSGSVNF-KNGNATFATVSKDNNGNINVKYDVANGDGLKIGDKKIYADTTLTV 1069  
QY 1472 GNRKVIYDSTDNKRYQAKNDGTVDKTEVAKDKLVAAQOTPDGTLAOMNYSKVINKEQVND 1531  
Db 1070 TGGKVSVPAGANSVNN-NKKLV-N-AEGLATALNLSWTAKAD-KYA-D-GESEGE-TDO 1122  
QY 1532 ANKKOINENDNAFVKLEKRAASDNKTKNAAYVGLDANAAQPLPFAAGDTGTAKKLGFT 1591  
Db 1123 -EYKADKVTFK-AGNKLKVKOSEKDFYSLQDTLGLTSLUGGANGRNDTGVINND 1180  
QY 1592 LTIKGGOTDINKLTLDNNGIGVAGTDFYKLAADLNLNLSVANGAKTID-DKGVSEFVDS 1650  
Db 1181 GLTITLALNG-AAAGTD-ASNG-NTIS-VTKDGSAGNKEITVNVKSLKTYKTOQTADT 1236  
QY 1651 GOKKANTPVLSANGDLGGKVISNVGKGTIDDAVVOULNEVRNLGLGNAGNNDADON 1710  
Db 1237 Q-DK-EFHAAVKNAEVE-FVGKNGATVSAKTDNNG-K-HYVITIDVAAKVGDLKEDTD 1291  
QY 1711 QVNIADIKKOPNGSSSNRTVYKAGTVLGGKNNDEKTLATGCVGVKDKGANDNLN 1770  
Db 1292 GKTKLKVNDTGNLLTVATIKASVAKGFENAVTTDATTAQST---NANENGKV-VYNG 1347  
QY 1771 VWKTKODGSK-KALLATYNAAGOTNVLNNPFAIDRINEQIRFEFHVNDGQEPVVG 1829  
Db 1348 SNGATETETKKKATVGVAKAINDAATV-K-VEN-DDSATIDDSPTDDGAND-ALKA 1403

QY 1830 RKGIDSSASGKHSVA-IGFQAKADEAAVAIGROTQAGNOSIALGDNAAOTGDOSIALGT 1888  
Db 1404 GDTLLKAKNKLKVRDKG-NITFALANDLSYKATVSDKLSLGTNGKNVIT-SDTKGL 1461  
QY 1889 GNVVAKHSGALIGDSTYKADNSYSVGNNGOFTDQTD-V-FGV-GNNITVETESNSVAL 1945  
Db 1462 NEAKDSKTGDDANIHNLNGTASTJDTLNLASGATYTLUG 1499  
QY 1946 GSNASISAGTHAGTOAKSDGT-AGTITTAGATGVKG 1982

RESULT 2  
ID Q48152 PRELIMINARY; PRT: 1098 AA.  
AC Q48152;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE ADHESIN (H1A).  
GN H1A.  
OS HAEMOPHILUS INFLUENZAE.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
OC PASTEURRELLACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NONTYPABLE STRAIN 11;  
RX MEDLINE: 96332658.  
RA BARENKAMP S.J., ST GEME J.W. III,  
RL MOL. MICROBIOL. 19:1215-1223(1996).  
DR EMBL: U38617; G1235666; -  
SQ SEQUENCE 1098 AA; 114100 MW; 0AD09C70 CRC32;

Query Match 3.38; Score 441; DB 2; Length 1098;  
Best Local Similarity 25.18; Pred. No. 1,37e+40;  
Matches 236; Conservative 223; Mismatches 407; Indels 73; Gaps 62;

Db 173 TATVSDTLTGGGAAGACTT-PRVNV-TSTTDGLFAKDAAGDITVHLNGIGSTLT 230  
QY 1243 TSKTSVITPAGENGITTKVKNKGVAVGIDQTKGLTTPRLTGVNNGKGIYIDSONGWT 1302  
Db 231 DTLVSPATHIDGSDSTHYTRAASI-KDVLNA-GMNKGY-KAG-STTGQSENVDFVHT 286  
QY 1303 ITGLSNTLAVTNDKSVRTTEOGKITIDEDKTRASIVDVLSAGFNLOGNEAVDFST 1362  
Db 287 YDVEEFLASDTEFTTVYDSKENGKRTVEYKIGAKTSVIEKDGKLTGKANKETNVDA 346  
QY 1363 YDVTNFDAGNATTAKYVDDTSKTSKVYDVAVDDTITVEYKDKL-GVYTTTLTSTGTA 1421  
Db 347 N--ATEDADGKGLYAKVIDAVNKTGMRIKTTDANGNGDFAT-V-ASGTNVTFASGN 402  
QY 1422 NKFLSNQATGADALVAKSDIVAHMLTSLGDIOTAKGASQANSAGYVADGNKVIYDSTD 1481  
Db 403 GTTATVNGDITTVYKADAVGDKLIDGDKIAADTTALTLYNDGKANNPKGVADVAST 462  
QY 1482 NKYYQAKN--DGVVDKTRK-VARD-KLVAAQOTPDGT-LAOMNYSKVIN-KEDVND-ANK 1534  
Db 463 DEKLVYAKGLVYALN-SLS-WTTTAAEDGTLGDNASBOEKAGADK-VTERA-GKNLK 518  
QY 1355 KQG-INEDNAFVKLEKRAASDNKTKNAAYVGLDANAAQPLPFAAGDTGTAKKLGFT 1593  
Db 519 VKQGANETYSLODALTGLTSLTGGNNGAKTEINKGLTTPANGAGANNANTISVTK 578  
QY 1594 IK--GGQ-T-D-TNKLTD-NNIGVAGTDFYKLAKD-LTINLSYNAGGTIKIDGVS- 1645  
Db 579 DGISAGOSYKAVNSGLAKFGDANFDTLSSANGLTKONDADAYKGLNDEKTDQTPV 638  
QY 1646 -FVDSGGA-KANTPVLSANGDLGGKVISNVGKGTIDTAA-N-VQOLNE-VRNILGIG 1700  
Db 639 VADNTAATVGLRGLGWTISADKTTGGSTEVHDVNRANVEKFGSNGINVSQKTVNGR 698  
QY 1701 NAGDNAD-GNOVVA-DIKKDPSSSSN-RVYIK-AGTV-L-GGKNNDEKTLATGCV 1754  
Db 699 EITTELAKGEVVSNEFTVETNGKETSLYKVGDKYSKEDIDLITGQPLKAGNTVAAK 758

(WL)

Distribution rights by Oxford Molecular Ltd

protein database search, using Smith-Waterman algorithm

Meda Oct 28 11:13:15 1998; MasPar time 135.60 Seconds 279.644 Milliseconds

...generalized ...

МННИУКВИЕНКАТГТЕМАВА.....

2:30 PM 150

165420 seqs, 49795644 residues

existing first 45 summaries

sptrembl6

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

mean 55.877; Variance 151.560; scale 0.369

by analysis of the total score distribution.

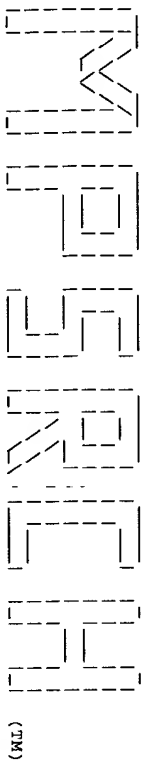
## SUMMARIES

## ALIGNMENTS

Query Match	Score	DB 2;	Length
3.9%;	518;		2353;
Best Local Similarity	33.68;	324	140

Matches 348; Conservative 380; Mismatches 701; Indels 109; Gaps 102;





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Oct 28 11:26:52 1998; MasPar time 66.30 Seconds  
518.153 Million cell updates/sec

Tabular output not generated.

Title: >US-08-968-685-10 / 09/8/93, 214 Seq ID #10  
Description: (1-2123) from US08968685.pep  
Perfect Score: 13368  
Sequence: 1 MNHKKVIFNKATGTFWAVA.....NGSADTQGHVGAAGVGFHF 2123

Scoring table: PAM 150  
Gap 11

Searched: 131922 seqs, 16180660 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq32  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29

Statistics: Mean 39.003; Variance 251.705; scale 0.155  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	8197	61.3	1992 20	W04505	Moraxella 200 kDa out	0.00e+00
2	518	3.9	2353 19	R99393	Haemophilus adhesion	6.07e-23
3	441	3.3	1098 19	R99392	Haemophilus adhesion	4.14e-18
4	424	3.2	1529 8	R41732	High molecular weight	7.56e-09
5	290	2.2	1598 27	W30281	Non-typeable Haemophi	5.00e-09
6	293	2.2	1601 27	W30282	Non-typeable Haemophi	8.68e-09
7	289	2.2	1338 8	R41731	High molecular weight	1.31e-08
8	286	2.0	43 28	W32274	M. catarrhalis outer	3.51e-07
9	262	2.0	1477 12	R63506	Haemophilus high mole	2.69e-06
10	247	1.8	1477 12	W30294	Non-typeable Haemophi	6.91e-06
11	240	1.8	1477 8	R41724	High molecular weight	6.91e-06
12	235	1.8	1822 5	R21775	Extracellular factor	3.02e-05
13	240	1.7	1477 8	R41728	High molecular weight	6.72e-05
14	229	1.7	1536 27	W30293	Non-typeable Haemophi	4.29e-04
15	223	1.6	1026 27	W37450	Caulobacter crescentu	1.49e-04
16	209	1.6	1536 12	R63505	Haemophilus high mole	1.49e-04
17	217	1.6	1536 8	R41723	High molecular weight	1.49e-04
18	217	1.6	1536 8	R41723	High molecular weight	1.49e-04

## ALIGNMENTS

RESULT ID	Score	Query Match	Length	DB ID	Description	Pred. No.
1	8197	61.3	1992 20	W04505	Moraxella 200 kDa out	0.00e+00
2	518	3.9	2353 19	R99393	Haemophilus adhesion	6.07e-23
3	441	3.3	1098 19	R99392	Haemophilus adhesion	4.14e-18
4	424	3.2	1529 8	R41732	High molecular weight	7.56e-09
5	290	2.2	1598 27	W30281	Non-typeable Haemophi	5.00e-09
6	293	2.2	1601 27	W30282	Non-typeable Haemophi	8.68e-09
7	289	2.2	1338 8	R41731	High molecular weight	1.31e-08
8	286	2.0	43 28	W32274	M. catarrhalis outer	3.51e-07
9	262	2.0	1477 12	R63506	Haemophilus high mole	2.69e-06
10	247	1.8	1477 12	W30294	Non-typeable Haemophi	6.91e-06
11	240	1.8	1477 8	R41724	High molecular weight	6.91e-06
12	235	1.8	1822 5	R21775	Extracellular factor	3.02e-05
13	240	1.7	1477 8	R41728	High molecular weight	6.72e-05
14	229	1.7	1536 27	W30293	Non-typeable Haemophi	4.29e-04
15	223	1.6	1026 27	W37450	Caulobacter crescentu	1.49e-04
16	209	1.6	1536 12	R63505	Haemophilus high mole	1.49e-04
17	217	1.6	1536 8	R41723	High molecular weight	1.49e-04
18	217	1.6	1536 8	R41723	High molecular weight	1.49e-04



24-MAR-1995: US-409995.  
 (VUSL-) UNIV ST LOUIS.  
 (UNIW) UNIV WASHINGTON.  
 Barenkamp SJ, St Geme JW.  
 WPI: 96-455364/45.  
 N-PSDB: T41476.  
 DR  
 PT Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in  
 vaccines against H. influenzae infection.  
 PS Claim 5: Page 66-73; 120pp; English.  
 CC Haemophilus adhesion protein HA2 (R99393) is associated with the  
 formation of surface fibrils involved in adhesion to various host  
 cells; it is also referred to hsf (Haemophilus surface fibrils).  
 CC Its amino acid sequence was deduced from a genomic DNA clone  
 (T41476) derived from Haemophilus influenzae type b strain C65.  
 CC Large quantities of recombinant HA2 can be produced in transformed  
 prokaryotic or eukaryotic host cells, for use in vaccines against  
 H. influenzae infection.  
 Sequence 2353 AA:

Query Match 3.98; Score 518; DB 19; Length 2353;  
 Best Local Similarity 22.6%; Pred. No. 6,07e-23;  
 Matches 348; Conservative 380; Mismatches 701; Indels 109; Gaps 102;

Db 43 lftvqanadedeelpvrita-pyl-sfhadkegtgekvtenungvifdnh-gvlyk 99  
 QY 473 lrtagvdddaatrgqlkknqthsalofrtvkrkvdnknndandskilivgknknpdgtq 532  
 Db 100 agatllkagnklikqntdestnaasfyslkkdlc-dltvateklcf-gangdkvdl 157  
 QY 533 vntlkrlk-gen-gv-dvttetngtvtfglnonngltvngnstlwnndlsykntnsnk-qio 588  
 Db 158 sdangllaktgngvnhlgnldetlpdavnlg-vlsssgfepnd--vekt-raatvkv 213  
 QY 589 vgaagilfttdltsnsk-pgagientrttrtrgicgpanntgslanrkprlptgngagkxl 647  
 Db 214 lnaagwnlkagactagavesvd-lvsay-nuveftgdknldvnlcakengktvckftr 271  
 QY 648 tnvosaitmpa-tnggoldfemrlstamteksgsamrtikdysvplrtfagotgpnvtrk 706  
 Db 272 kts-vikekdqklftgkendntkvtantatndtdegnqlvtaekavdaenagkwtvkt 330  
 QY 707 klgeilkrvkgkrttadtltknnigvados-tdnsltrk-laktlslddanvntktrlasdk 764  
 Db 331 tangnggfataagvntvtesgdtasvckdctngit-vky-dak-vg-dgktdsd 386  
 QY 765 vtvdsngnntakong-dlfrsknongatpantnsrtiivgdikftdnnngalldogttrtk 823  
 Db 387 kklv-a-dtalvtgkva-eiak-ed-d-kklvnaqd-lvtaaln-lswkakaeadt 438  
 QY 824 dkvfgakogsdkskpfyldnkdlkvegttngingnsgaigtstltdatnatgth 883  
 Db 439 dgalglsidgdevkagetv-ftkagknllvkqdganf--cyslqdalqtlstlsgttn 495  
 QY 884 vtqlgltvdstdktraasidvlnagennlknngdardfvtvdyvding-natt-akvty 941  
 Db 496 gndaktv--inkdgliti--cpa-gunggtcgtntsvckdjkagknklt-nvasql-ra 548  
 QY 942 dgkaskvaydvndgttthltgdngrknqigvktttlvttdak-gkainsvnsddka 1000  
 Db 549 ydaa-nf-dvlnsatdlnhvedaykqllnhnekanqplv-tcstaavgdtrkxgw 605  
 QY 1001 linkadladvnltagetrtktadpaltotfvykvrkkgdddnadrtivgkgnaknq 1060  
 Db 606 vsvtk-ngtkesngvkvadevlfctgagaavtskengkhltvsvaetkdcgldkx 664  
 QY 1061 vntlkkgknngldiotnkcdvtvfgintogslagannntltnngsltkrtatnqioyva 1120  
 Db 665 dtkll-kvndgntnhtvngngfayvkggfetvctcatadagkytvvkaataandkxv 723  
 QY 1121 dgvrfaaknnngvavagld-gttrtrtrbdtgagtn-gslkspkphls-kdgi-nagkri 1176  
 Db 724 atvk--dvatainaatfvktenlttsided-optsdgdalagkdtlftkagknlkv 780

QY 1177 tntogseiaonsdnavtgckiydtktelemnkststaktakonslbf-svadeognnftvs 1235  
 Db 781 rdgknltfdalakhlevkt-akvsdlt-ignttpggtatpaxvntstaddlfnafeta 838  
 QY 1236 -np-yssydrsktsdvtvtefagengittkvngkvrvvgldotglttppkltvgnnnkglv 1293  
 Db 839 -daegskvnylkgfatltle-psa-ga-kshvndlvdatkksnaasiedviraagwimg 894  
 QY 1294 idsonqont-ittgslntlavntwtdksgsvrtteogkllrkedxtkraasivdvlsagfmlqg 1352  
 Db 895 ngmvdvayatyvntfddstglttvtvlgkad-gkyga-dvrlgaktvsklthngklftg 952  
 QY 1353 ngeandfvtstidyvnpadgnatrtakvtyddtstkskvdydvndvtdtievkdkkllcv-kt 1411  
 Db 953 kdlkdaangatvseadgdtdgtgvtvktvldavnk-sgvrvtgegat-aetgatavag 1010  
 QY 1412 tltstgckarkfalsnogatdalyknsdylahnltsldioyakaqsqnsagvydad 1471  
 Db 1011 naetvsgtsvnt-kngatattvskdnglnvkvdyvngdglkgdsklvadttltcv 1069  
 QY 1472 gknvtydstdknyqakndgtvdktevakdklvaqatpdtlaqmnvksvlnkeovnd 1531  
 Db 1070 tggkvsvpagansvnn-nkklv-n-aeglatelnlswtakad-kya-d-gesege-tdq 1122  
 QY 1532 ankkoqindnafaevglekaasdnkrtknaavtvgdlnavotplrtfagdtgtrakkltget 1591  
 Db 1123 -evkagdkvtfk-egknllvkqsekdfyslqdtlgtlslgtangrntdgtvlnkd 1180  
 QY 1592 lrtkgqotlntkltdnnglvavagtdgftvklakdltnlnsvnagatkd-id-dkgsfvds 1650  
 Db 1181 gltltlang-aaagtd-aang-nlis-vtkdglasgnkeltvksaalktytdntadet 1236  
 QY 1651 gqkamtptvlsangldlkgkvisnvgkgtkdndaanvoolevrrllglgnagndmndn 1710  
 Db 1237 q-dk-ethaavkhanave-fvngatvsaaktang-k-hvtldvaeakvgdglektd 1291  
 QY 1711 qvntadtkkdpnsgsssnrvtlragvllgkgnndpklatgvgvgydkgnanagldsn 1770  
 Db 1292 gklklykndidgmllvtatkaasvakefnatvdatatagtt--nanegkv-vvkv 1347  
 QY 1771 vwtvtoqkdgsk-kallatvnaacotvnltnnpaeldrinedgtrfeyhndnqepvvg 1829  
 Db 1348 ngatatecdkkyatvgtvkaakdaatfv-k-ven-dsatsldspddgand-alka 1403  
 QY 1830 rngidssasghsva-igfoakadgpaavalgroqagnsiaicgnaoatdosiait 1888  
 Db 1404 gdtllkagknllkxrdgk-nltfalandsvksatvsklsigtngvnt-satxgl 1461  
 QY 1888 gnvavagksagilgopstvykadmnsysvgnnnoftdatqtd-v-fgv-gnnltvtesvavl 1945  
 Db 1462 nfakasktgddanlhnglaetlcllsgatnlnlg 1499  
 QY 1946 gsnsaisagthagtakksdgt-agttrttagatgtvng 1982

# RESULT 3

AC R99392 standard; protein; 1098 AA.  
 ID R99392; AC R99392; (first entry)  
 DT 15-JAN-1997  
 DE Haemophilus adhesion protein HA1.  
 KW Haemophilus adhesion protein; HA1; hia protein; vaccine.  
 OS Haemophilus influenzae non-typable strain 11.  
 PN W09630519-A1.  
 PD 03-OCT-1996.  
 PE 22-MAR-1996; U04031.  
 PR 24-MAR-1995; US-409995.  
 PA (VUSL-) UNIV ST LOUIS.  
 PA (UNIW) UNIV WASHINGTON.  
 PI Barenkamp SJ, St Geme JW;  
 WPI: 96-455364/45.  
 DR N-PSDB: T41475.  
 PT Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in  
 vaccines against H. influenzae infection.



KM HMM: high molecular weight protein; virus; vaccine; influenza;  
 KM epitope; immunity; haemophilus influenzae.  
 OS Haemophilus influenzae.  
 PN W09316090-A.  
 PD 30-SEP-1993.  
 PR 16-MAR-1993; U02166.  
 PR 16-MAR-1992; GB-005704.  
 PA (BARE/) BARENKAMP S J.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PI Barenkamp SJ.  
 DR WPI: 93-320683/40.  
 DR N-PSDB: Q49511.  
 PT High molecular weight surface proteins - of non-typeable  
 PT haemophilus which exhibit immunogenic properties  
 PS Claim 6: Figure 10: 100pp: English.  
 CC The isolation and purification of the high molecular weight protein  
 CC enables the identification of the major protective epitopes of the  
 CC protein by conventional epitope mapping. These epitopes can then be  
 CC synthesised using standard techniques and incorporated into fully  
 CC synthetic or recombinant vaccines.  
 SO Sequence 1529 AA;

Query Match 2.2%; Score 290; DB 8; Length 1529;  
 Best Local Similarity 18.9%; Pred. No. 7,566-09;  
 Matches 145; Conservative 212; Mismatches 370; Indels 39; Gaps 37;

DB 686 nltsraagimdsi-n-ltsgldfstehtn-nnafeikkltnatgnsfslqtkds 742  
 QY 723 DLTNNIGVADSTNSLTVKAKTLSDLDANVTTLASDKVTVYDSCNNNTAKLNGDLT 782  
 DB 743 fyneyskhalnshnltilgnvltgennssstigninit-nkanvltgadit-snsntg 800  
 QY 743 FSKQNTG-ATPAINKKTIYVDGLKFTDNNNGIALDGTYYIKDKYGFKAQKODSLDKSPY 841  
 DB 801 lkrtltlgnisvegn-lsltg-ananlygnislaedstfkkeasdnitlgtfnngta 858  
 QY 842 LDKDKLKGEVEITTINGINAGSKATITGLSNTLTDTATNAT-TGHVYQ-LGIYDSTDKTAA 899  
 DB 859 ni-nl-kqgv-vklagd--link-gglnlttnasgtqkltlgnitlnekgdlnknka 911  
 QY 900 SIGDVINAGFNKLNNDADDFVSTVDVFINGNATTAFTVDGKASKAVAYDVNDGTTI 959  
 DB 912 deldqig-n-isqkegnltiesdkvntlnqtlkagveggssdsse-aenan-ltltqtk 967  
 QY 960 HLTGADGNKNOIGVKTTLTKTDAKGDKAINFSVNSGDKALINKADIANNTLAGETR 1019  
 DB 968 elklagdlinsfknkeitakngsdltignasg-nadaky-ttd-kvks-kistdgh 1023  
 QY 1020 NTGKTADTLQTFQVKKVKEKNGDDDDADDTIVGKDAKTNQVNTLKLKKNGLDIQTND 1079  
 DB 1024 n-vtltnsevktsngssnagndstglitlsakdvtnnvtshkltlnisaaagvttkegt 1082  
 QY 1080 GVTYFGINQSGLKAGNNNTLNNGLSIKNTAGNEQIQVAGGVKFAKNNVGVGAGIDG 1139  
 DB 1083 tlnatgsvevtngtlt-kgnit-sgnvtvcatenltvtenavlnatsgtvntsktgd 1140  
 QY 1140 TTRITRDEIGFAGTNGSLKSKPHLSKDGINAGKRTITNIQSEIIONSDAVTGKTIYD 1199  
 DB 1141 ikgsiest-sgnvntcasg-nltkvnitqgdvtvtadagallttagstisattgnait 1198  
 QY 1200 LKTELEKISTAKTAKONSHIESFVADDEGONNTVSNPYSYDTSKTSOVITPAGENGIT 1259  
 DB 1199 tkgtgdlnkveasssvtlyatgaclavagnisgnvtltedsgklst-vgstingtnsv 1257  
 QY 1260 TKV-N-KGVVRVIGIDOTKGLT-PLKLVGNNGKGIYIIONQONNTITGSLNLANVTND 1316  
 DB 1258 tltsgsdigltsgnvtvntastgdltinsakveekgaatltesgklttqtsait 1317  
 QY 1317 KGSVRTEGKRIKDEKTRPAASIVD-LSAGFNLOGNGAVPVFSTYDVNADGNATT 1375  
 DB 1318 snngtltakdsiaagninaavntltgt-lttigdsinatstgltlnakdaldga 1376  
 QY 1376 AKYIYDD-TKRTSKVYVDVAVDOTTIEVKDKKIGVKTTLTSTGTGANKFALSNOATGDA 1434

DB 1377 asgdtvtnatnassgnv-taktssvntigdlntinglnisen 1421  
 QY 1435 LVKASDIYAHNLNTL-SGDIQTAKGASQANSSAGYADADGNKVIYDS 1479

RESULT 6  
 ID W30291 standard; Protein: 1598 AA.

AC W30291;  
 DT 14-APR-1998 (first entry)  
 DE Non-typeable Haemophilus high mol.wt. surface protein HMM3.  
 KW Non-typeable Haemophilus; high molecular weight surface protein;  
 KM HMM3; immunogen; vaccine; otitis media.  
 OS Haemophilus influenzae strain 5.

FT Key Location/Qualifiers  
 FT Misc-difference 113  
 FT /note="encoded by GTC"  
 FT /note="encoded by TGT"

PN W09316091-A1.  
 PD 09-OCT-1997.  
 PR 01-APR-1997; U04707.  
 PR 01-APR-1996; US-617697.  
 PA (BARE/) BARENKAMP S J.  
 PI Barenkamp SJ.  
 DR WPI: 97-503038/46.  
 DR N-PSDB: T90992.

PT High molecular weight proteins of non-typeable Haemophilus  
 PT influenzae - useful for vaccine production  
 PS Claim 1: Page 93-97; 183pp: English.  
 CC This protein comprises the high molecular weight surface protein  
 CC HMM3 (125 kDa) of non-typeable Haemophilus influenzae strain 5 that  
 CC has the immunological ability to protect against disease caused by  
 CC a non-typeable Haemophilus strain and is characterised by at least  
 CC one surface-exposed B-cell epitope that is recognised by monoclonal  
 CC antibody AbC. The HMM3 amino acid sequence was deduced from an  
 CC isolated hmw3 gene (see T90992). HMM1 (see W30293), HMM2 (see  
 CC W30294) and HMM4 (see W30292) have also been identified. A  
 CC conjugate comprising HMM3 linked to an antigen, hapten or  
 CC polysaccharide, and a synthetic peptide of 6-150 amino acids  
 CC corresponding to at least protective epitope of HMM3 are also  
 CC claimed. HMM proteins, conjugates and peptides can be used in  
 CC vaccines, as immunogens for preparation of antibodies and as  
 CC antigens for detection of these antibodies.  
 SO Sequence 1598 AA.

Query Match 2.2%; Score 293; DB 27; Length 1598;  
 Best Local Similarity 19.0%; Pred. No. 5,006-09;  
 Matches 158; Conservative 237; Mismatches 393; Indels 44; Gaps 40;

DB 682 lsldstgsstgspisina-elngitfnkatfnlaqstansfsikasmpfknsanyaltn 740  
 QY 669 LSTANTEKSGSAATIKDLYNLSQVPLFFAG-DIGPNVTKKLGLLAKVKGKKTAD-DLTK 726  
 DB 741 edisv--srggslnfklnassnigtprv-ilk-sgnfivsgsgstclnlaegstetafs 795  
 QY 727 NNIGVADSTNSLTVLAKTILSDLDVNTKTLTASDKVTVYDSCNNNTAKLNGDLT 784  
 DB 796 lendlnatgntltqvgvgtstsvnkyaaknltfksgnltffgsqkateltkgnvtl 855  
 QY 785 KONTCARPATNSK-TIYGVDELKFTDNNNGIALDGTYYIKDKYGFKAQKODSLD-KSKPYL 842  
 DB 856 nknatrlrganfaenksplnagvlnnqnltnagsltnaglnltvskganlqatnvt 915  
 QY 843 DKD-K-LKVEVELTITNGINAGSKAIT-G-LSMT-LTD-ATNATGHTVTOIGIVDSTD 894  
 DB 916 fnvsgfnnaganslatargakfkdlnt-sslnitnsdltlyrtliknisksgdln 974  
 QY 895 KTRASIGDVINAGFNKLNNDGDA-KDFVSTVDVFINGNATTAFTVDGKASKAVAYDVN 933  
 DB 975 ldkksdeldqig-n-isqkegnltiesdkvntlnqtlkagveggssdsse-aenan- 1030  
 QY 954 VDGTTIHLTGADGNKNOIGVKTTLTKTDAGDKAINEFSVNSGDKALINKADIANNTLAGETR 1013

Db 1031 lctqtkelkylagdlinsgfnkkaetkngsdltlignasgg-nadakkv-tfd-kvkds-k 1086  
 QY 1014 lgeirntkgtadlqotqvkrkknegddndadtltvgkdkaktnovwtlklkngld 1073  
 Db 1087 lstdghn-vtlasevktngssnagnadstglttsakdctvnnvtskhlisaaagv 1145  
 QY 1074 lctnndgvtvtfintnosgkacnnttlnnngslskntkgnedioqgaovkfrakvngv 1133  
 Db 1146 tkkgeltlnatgsvetnaglti-kgnlt-sgnvltatenlvtetenavlnatsgtvni 1203  
 QY 1134 gagiogtrrtfdeigfagctngslskpshlskdginagkktitngsgelaaqnsndavt 1193  
 Db 1204 sktgdligjstest-sgnvntasg-nlksvntlgdvtvtadagalttaagstisatc 1261  
 QY 1194 gkrtldkteleknsktstaktqnslhefsvadeqgnnftvsnpyssvtdstktspvtrfa 1253  
 Db 1262 gnaaitlktgdlngkveessgsvltatgatlavngisgnvltladsqkltst-vgsti 1320  
 QY 1254 gencitfky-n-kgyvrvigldotkgltr-pklvgnnnngkgtvldsqngonttltgslwtl 1310  
 Db 1321 ngtnsvtssgsgdlegltisgnvntasgdlitignsakveakngaatlaesgklttg 1380  
 QY 1311 anvtndksvrtteogkrlkdedktrasiydv-lsagfnlqngnagvdfvstydtnfa 1369  
 Db 1381 tgsstisngqtlktakdsiagnnaavntltgt-lttgsklnatstgltlnakd 1439  
 QY 1370 dcnattavtydd-tskskvyvdyvnddttevkdkklyvktttltstgtgafalsn 1428  
 Db 1440 akldgaasgdtlvnatnasggnv-taktssvntlgdltnglnlisen 1490  
 QY 1429 oatgdalvkaadivahlntl-sgdiofakgasqanssagvydadgnkviyds 1479

RESULT 7  
 ID W30292 standard; Protein; 1601 AA.  
 AC W30292;  
 DT 14-APR-1998 (first entry)  
 DE Non-typable Haemophilus high mol. wt. surface protein HMW4.  
 KW Non-typable Haemophilus; high molecular weight surface protein;  
 KM HMW4; immunogen; vaccine; oltis media.  
 OS Haemophilus influenzae strain 5.  
 FT Key Location/Qualifiers  
 FT Misc\_difference 372  
 FT /note- "encoded by Tcr"  
 FT /note- "encoded by AAT"  
 FT Misc\_difference 400  
 PN W09736914-A1.  
 PD 09-OCT-1997.  
 PF 01-APR-1997; 004707.  
 (BARE/) BARENKAMP S J.  
 Barenkamp SJ;  
 DR WPI: 97-503038/46.  
 DR N-PSDB; T90993.  
 PT High molecular weight proteins of non-typable Haemophilus  
 PT Influenzae - useful for vaccine production  
 PS Claim 1; Page 97-102; 183BP; English.  
 CC This protein comprises the high molecular weight surface protein  
 CC HMW4 (123 kDa) of non-typable Haemophilus influenzae strain 5 that  
 CC has the immunological ability to protect against disease caused by  
 CC a non-typable Haemophilus strain and is characterised by at least  
 CC one surface-exposed B-cell epitope that is recognised by monoclonal  
 CC antibody A56. The HMW4 amino acid sequence was deduced from an  
 CC isolated HMW4 gene (see T90993). HMW1 (see W30293), HMW2 (see  
 CC W30294) and HMW3 (see W30291) have also been identified. A  
 CC conjugate comprising HMW4 linked to an antigen, hapten or  
 CC polysaccharide, and a synthetic peptide of 6-150 amino acids  
 CC corresponding to at least protective epitope of HMW4 are also  
 CC claimed. HMW proteins, conjugates and peptides can be used in  
 CC vaccines, as immunogens for preparation of antibodies and as  
 CC antigens for detection of these antibodies.  
 SQ Sequence 1601 AA.

Query Match 2.2%; Score 289; DB 27; Length 1601;  
 Best Local Similarity 18.9%; Pred. No. 8.68e-09;  
 Matches 145; Conservative 217; Mismatches 364; Indels 41; Gaps 39;

Db 758 nltsraaimdsi-n-ltggldfsitsmr-nsafelkkltnatgnsfslqtkds 814  
 QY 723 dlfrknnigvvaadstnslfvyklaktlsddavntklrtlsdsvtytdsgntakltongltr 782  
 Db 815 fyneyskhaanshnlitlggnvltlgensssisignlilt-nkanvltgadl-snsntg 872  
 QY 783 fskontg-atparysktitygvdlkftdnngalldgtyttrtdkxgfkqdsldkskpy 841  
 Db 873 lkrtrtlgnisvegn-laltg-anaavlgnsiaedstfkeasdnltgtftngta 930  
 QY 842 ldrdklkvgevelttnnginagkaltlglstltdatna-tchvqo-lgtydstktktra 899  
 Db 931 nl-nl-kgyv-vklgd---link-gglnltuasgqtkltlignltnekgdlnlka 983  
 QY 900 stldvlnagfnnknngdardfvttydtdvfingnatfakvtydkaskavaydvndgvti 959  
 Db 984 daeiqig-n-lsqkegnltsdskvnltnqtlitkagvegrtsdase-aenan-ltqtk 1039  
 QY 960 hltagdgnknqogvktttltktadagdkalnsvnsdgdkaalnkdlnlntlagelr 1019  
 Db 1040 eklagdlinsgfnkkaetkngsdltlignasgg-nadakkv-tfd-kvkds-kistdgh 1095  
 QY 1020 ntfgtrdlatqtfvqkrkknegddndadtltvgkdkaktnovwtlklkngldioinkd 1079  
 Db 1096 n-vtlasevktngssnagnadstglttsakdctvnnvtskhlisaaagvntkeg 1153  
 QY 1080 gvtvtfintnosgkacnnttlnnngslskntkgnedioqgaovkfrakvngvvaagid 1138  
 Db 1154 tlnatgsvetnaglti-kgnlt-sgnvltatenlvtetenavlnatgsvtnlsktg 1211  
 QY 1139 gtrtrtrtdelgfragctngslskpshlskdginagkktitngsgelaaqnsndavt 1198  
 Db 1212 dkggiest-sgnvntasg-nlksvntlgdvtvtadagalttaagstisatgna 1269  
 QY 1199 dlkteleknkisttaktqnslhefsvadeqgnnftvsnpyssvtdstktspvtrfa 1258  
 Db 1270 tktgdligkveessgsvltatgatlavngisgnvltladsqkltst-vgsting 1328  
 QY 1259 ttrkv-n-kgyvrvigldotkgltr-pklvgnnnngkgtvldsqngonttltgslwtl 1315  
 Db 1329 vltssgsgdlegltisgnvntasgdlitignsakveakngaatlaesgklttgts 1388  
 QY 1316 dkgsvrtteogkrlkdedktrasiydv-lsagfnlqngnagvdfvstydtnfa 1374  
 Db 1389 tsngqtlktakdsiagnnaavntltgt-lttgsklnatstgltlnakd 1447  
 QY 1375 takvtydd-tskskvyvdyvnddttevkdkklyvktttltstgtgafalsn 1428  
 Db 1448 aasgdlrvnatnasggnv-taktssvntlgdltnglnlisen 1493  
 QY 1434 alvrasdlvahlntl-sgdiofakgasqanssagvydadgnkviyds 1479

RESULT 8  
 ID R41731 standard; Protein; 1338 AA.  
 AC R41731;  
 DT 26-APR-1994 (first entry)  
 DE High molecular weight protein 3 (HMW3).  
 KW HMW; high molecular weight protein; virus; vaccine; influenza;  
 KW epitope; immunity; haemophilus influenzae.  
 OS Haemophilus influenzae.  
 PN W09319090-A.  
 PD 30-SEP-1993.  
 PF 16-MAR-1993; 002166.  
 PR 16-MAR-1992; GB-005704.  
 PA (BARE/) BARENKAMP S J.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PI Barenkamp SJ;



DR WPI: 93-320683/40.  
 DR N-PSDB; Q49510.  
 PT High molecular weight surface proteins - of non-typeable  
 PT haemophilus which exhibit immunogenic properties  
 PS Claim 5; Figure 10; 100bp; English.  
 CC The isolation and purification of the high molecular weight protein  
 CC enables the identification of the major protective epitopes of the  
 CC protein by conventional epitope mapping. These epitopes can then be  
 CC synthesised using standard techniques and incorporated into fully  
 CC synthetic or recombinant vaccines.  
 SQ Sequence 1338 AA;

Query Match 2.1%; Score 286; DB 8; Length 1338;  
 Best Local Similarity 18.9%; Pred. No. 1.31e-08;  
 Matches 157; Conservative 238; Mismatches 393; Indels 44; Gaps 40;

422 1s1dsgstgspisira-elnqitfnkatfnlaagstanfsikaspfksanayalfn 480  
 669 LSTANTKESGAARITKOLYNISQVPLFAG-DTGPNTKRLGELKYGKGTAD-DLTK 726  
 481 edisv---sggsvnfknasssnitqpyv-ilk-sqnfsvsgstnlkaegstetafs 535  
 727 NNIGVAVDSTDNLSLVKLAKTLSDLDAVNTKTLASDKVYDVGNNMRA-KLQNG-DLTF 784  
 536 iendlnatognltirvgvgtsvrnkyaaknltffkgntffgskatteikgnvtl 595  
 785 KQNTGAPPAIRSK-TTGVDSGLKFTDNNGLALDGTYYITKDKGFAGQDSLD-KSKPYL 842  
 596 nknatnrganfaenksplnagvlnnqnltagsslnnaglnlvskaqlaqlnnt 655  
 843 DKD-K--LKVEVELTNGINAGKAIT-G-LSMT--LTD-ATNATGHTVQIGIVSTD 894  
 656 fnvgsfmgnaasnsiargakfkdnnt-slnitnsdtyrllkgnnsqgdln 714  
 895 KTRASISDVNAGFNKNNGDA-KDFVSTYDVTDFNGNATKATKYDGAASKVAVDVN 953  
 715 iidskdeieigv-n-isokegnltissdkvntnqtlkagvrgsdsse-aenan- 770  
 954 VDGTTHTLGTADGNKNQIGVATTTTLTKDAKGAINEFVNSGDKALINAKDADNMLNT 1013  
 771 ltiqtkelagadlnisgfnkaetknsgdltignassgnadakkv-tfd-kvks-k 826  
 1014 LAGELRNKKGADNALQFFQYKVKRENGDDNDADTIVGADATNQVNTKLKGKNGLD 1073  
 827 istgghn-vltnsevktsngssnagndstgltlisakdvtnnvtshklnisaaagnv 885  
 1074 IQTNKDGTVTFGIMTOSGLKAGNMTLNNGLSIKNTAGNFOIGADGVKFAKVNNGVV 1133  
 886 ttkegtlnatgsvevtaqngtl-kgntt-sqnvvtatealnvttenavnatsgvtvl 943  
 1134 GAGIDGTRTRTDEIGFAGTGSLSDKSPHLSKDGINNGKKTITNIOGGEIQAQNSNDAYT 1193  
 944 stktgdikgilest-sgnvnltasg-nlkyvnltgqvtatadagalttagstiaat 1001  
 1194 GCKIYDTELEENKSSAKTAQNSLHEFVADDEGNNFTVSNPSSYSDTSKSDVTIFA 1253  
 1002 gnanitktgdnkvsssgsvtlvatgatlavagnisgnvtltadagkltst-vgsti 1060  
 1254 GENGITTRV-N-KGVVVRGIDOTKGLT-PRITLVNNGNKGIIVDISQNGQMTITGLSNTL 1310  
 1061 ngtnsvtsssggdiegtlsgntvntastgdlitignsakveakngaatllaesgklttg 1120  
 1311 ANVTNDKGSVTRTEGKTIKDEKTRASIVDV-USAGFNQNGEAEVFTYTDYVNEA 1369  
 1121 tgsstsgngtlttakdsiagnnaavlnltgt-1tttgdsaknatagtltnakd 1179  
 1370 DGNATTAATYVD-TSKSKVYVDVNDVTLEVADKRLGKTKTTLTSTG@CANMFALSN 1428  
 1180 akldgaasgdtvtnatnasggnv-taktssvnltygdlnltnglnlsen 1230  
 1429 QATGALVAKASIVAHMLTL-SGDIOTAKGASQANSSAGYDADGNKVLYDS 1479

RESULT 9  
 ID W32274 standard; peptide; 43 AA.

AC W32274;  
 DT 08-MAY-1998 (first entry)  
 DE M. catarrhalis outer membrane protein (OMP)-106 peptide fragment 1.  
 KM Outer membrane protein-106; OMP106; vaccine; immune response;  
 KW cytotoxic antibody; Moraxella catarrhalis.  
 OS Moraxella catarrhalis.  
 PN M09741731-A1.  
 PD 13-NOV-1997.  
 PF 28-APR-1997; U07679.  
 PR 03-MAY-1996; US-642712.  
 PA (ANTE-) ANTEX BIOLOGICS INC.  
 PI Plosila L. Tucker K;  
 DR WPI: 97-558601/51.  
 PT Outer membrane protein, OMP106, of Moraxella catarrhalis - used in  
 PT vaccines for producing immune responses against M. catarrhalis  
 PS Claim 9; Page 23; 78pp; English.  
 CC This is a peptide fragment of a novel outer membrane protein-106 (OMP106)  
 CC polypeptide. The OMP106 is an outer membrane polypeptide of Moraxella  
 CC catarrhalis, an haemagglutinating cultivar. The peptide fragment can  
 CC specifically bind to an antibody that binds the OMP106 polypeptide. The  
 CC antibody is a cytotoxic antibody which mediates complement killing of  
 CC M. catarrhalis. The OMP106 polypeptide, and its peptide fragments can be  
 CC used in vaccines and antigenic compositions. They can also be used for  
 CC producing an immune response in an animal against M. catarrhalis.  
 SQ Sequence 43 AA;

Query Match 2.0%; Score 262; DB 28; Length 43;  
 Best Local Similarity 100.0%; Pred. No. 3.51e-07;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 1giseadgkqganarqgksiaidgialsgsaiadnkxv 43  
 69 IGISEADGKQGANARQGSIAIDGIALSGSQAIDGNKIV 111

RESULT 10  
 ID R63506 standard; Protein; 1477 AA.

AC R63506;  
 DT 25-JUN-1995 (first entry)  
 DE Haemophilus high molecular weight protein HMW2.  
 KM High molecular weight protein; HMW2; protective vaccine; otitis;  
 KW sinusitis; bronchitis; Hib; ss.  
 OS Haemophilus.  
 PN W09421290-A.  
 PD 29-SEP-1994.  
 PF 15-MAR-1994; U02550.  
 PR 16-MAR-1993; US-038682.  
 PA (BARE/) BARENKAMP S J.  
 PA (SGEK/) ST GEORGE J W.  
 PI Barenkamp SJ, St GEORGE JW;  
 DR WPI: 94-316665/39.  
 DT O-PSDB; Q72294.  
 PT New immunogenic high mol. wt. proteins of non typeable  
 PT Haemophilus - useful in protective vaccines  
 PS Claim 3; Page 36; 127pp; English.  
 CC The HMW2 protein encoded by this sequence is useful in a vaccine to  
 CC protect against disease caused by non-typeable Haemophilus which are  
 CC not controlled by H. influenzae type b (Hib) vaccines. The encoded  
 CC protein can also be used as a carrier for protective Hib  
 CC polysaccharide (in a conjugate vaccine against meningitis) or for  
 CC other antigens, haptens, etc.  
 SQ Sequence 1477 AA;

Query Match 1.8%; Score 247; DB 12; Length 1477;  
 Best Local Similarity 19.3%; Pred. No. 2.69e-06;  
 Matches 155; Conservative 228; Mismatches 373; Indels 48; Gaps 46;

DB 584 daanakiavagvtlt-tgegkdfirannvel-ngtykglnissvmltnlsgtlnisgn 641  
 1255 ENGITTKVNGVVRGVIDOTKGLTTPKLTGNNNGKGI-VIDSQNG-QMTITGLSNTLAN 1312





QY 1371 GNATTAKVYDDTSKTSKVYDVNVDDFTIEVKDKLGVK-TTTLSTGTGANKFL-SN 1428  
 Db hsgygaeklmseinsnganftlnshvr-gddafkinkdlt-lnatnsfslqtkdfl 815  
 QY 1429 QATGDALYKASDI-VAH-LN-TLSGDIOTAKGASQANSAGYVDADNKIYDSTDNKY 1485  
 Db 816 dyaenalnstynisi--lgnv--tlggqssssitnltcleaanvtleannpqn 872  
 QY 1486 QAKNDGTVDKTEVAKDKLVAQAOTPDGTLAQMNVKSYINKEQVNDKROGKINEDNAFV 1545  
 Db 873 rdryikgsllyvngslst-ge-nadikgnltis-esatfgykrctlnltgnfngt- 928  
 QY 1546 KG-LEKASDNKTKNAATVGDNLNAVAOTPLTFEGDITTKKIGEHLITIGSGTIDNKL 1604  
 Db 929 aenitgvtvklg-nvtdgdl-nlt-hakrngsllggllnkkslnitdsndae- 984  
 QY 1605 TDNNTGVAVAGDGFYVLAKLKDLTINLSVNAVAGTKIDKGVSEFVDSGQAKANTPVLASNG 1664  
 Db 985 lqiggnisqkegnltissdklnltkqtlkqldgedssdsatnaltkclkted 1044  
 QY 1665 LDIGKRVISNGKGTCTPDANV-QOLNEVRNLGLGNAGNDMADGN-QVNIADIK-KDP 1721  
 Db 1045 lsiqfmaeeltakgrliltignsndgnsgeakvtfnvkdqklsadghnvtlnskvk 1104  
 QY 1722 NSGSSSNRT-VI-KAGTVLG-GKGNNDTEKLATGCVQYGVKDGANADLSNVWKTQ-K 1777  
 Db 1105 tessngres-nsdndtg-lilt-aknve-vnkdltslktynitasekvttagsltinat 1160  
 QY 1778 DSKKALATYVAAQOTVYLTNNPAEIDRIEKGIRFEHVNDGNOEPVVOGRNGIDSSA 1837  
 Db 1161 ngkasitck-tgdlsqtlsgntsvsaqgdlttksgskleaksgeantstgtl-ggtl 1228  
 QY 1838 SKRHSVAIFQKADGEAVALIGROTAQONOSIAIGDNAQA-TGQOSTALIGGNVAAKH 1896  
 Db 1219 sgntv-vranagd-ltyvngaeinatagaeltatgnltceagsstskqvdllaq 1276  
 QY 1897 SGATGDPSTVAKADNSYVNNNOFTDATOTDVFVGNNITTFESNSVALGSSAISAGTH 1956  
 Db 1277 nsiagslnaanvltntgtlittvagsdikatsgtlvinakda--klngdsqdslevna 1334  
 QY 1957 AGTQAKSDGTAGTATTAGATGTVKGFAGQATAVNAVSGASGERRIQNVAAGEVSAST 2016  
 Db 1335 vnaagsgsvtaatsssvnlgtln 1358  
 QY 2017 DAVNGS-OLYKATOGIANATNELD 2039

RESULT 12  
 ID R41724 standard; Protein: 1477 AA.  
 AC R41724;  
 DT 26-APR-1994 (first entry)  
 DE High molecular weight protein 2 (HMM2).  
 KW HMM: high molecular weight protein; virus; vaccine; influenza;  
 KW epitope; immunity; haemophilus influenzae.  
 OS haemophilus influenzae.  
 PN W09319090-A.  
 PD 30-SEP-1993.  
 PF 16-MAR-1993; 002166.  
 PR 16-MAR-1992; GB-005704.  
 PA (BARE/) BARENKAMP S J.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PI Barenkamp S J.  
 DR WPI: 93-320683/40.  
 DR N-PSDB: Q49507.  
 PT High molecular weight surface proteins - of non-typeable  
 PT haemophilus which exhibit immunogenic properties  
 PS Claim 4: Figure 4: 100bp. English.  
 CC The isolation and purification of the high molecular weight protein  
 CC enables the identification of the major protective epitopes of the  
 CC protein by conventional epitope mapping. These epitopes can then be  
 CC synthesised using standard techniques and incorporated into fully  
 CC synthetic or recombinant vaccines.  
 SQ Sequence 1477 AA.

Query Match 1.8%; Score 235; DB 8; Length 1477;  
 Best Local Similarity 18.7%; Pred No. 1,356-05;  
 Matches 284; Conservative 438; Mismatches 666; Indels 110; Gaps 99;

Db 1 mnhklyrflksrlnalvavselargcdhstekysepkm-kvrlhalpjsam-lslg 58  
 QY 1 mnhklyrflksrlnalvavselargcdhstekysepkm-kvrlhalpjsam-lslg 58  
 Db 59 vrsipgsvlaagldgmavhytatamgvgnkltlnsdalimwqfildgmewqfildg 118  
 QY 59 AT-LINGSAYA-GIGISEADGKGGANARGDKSIAIGDIAQAGSOSIAGDKYH--NS 114  
 Db 119 nmnsavfnrtvnglsqklg-lldnggvflinpditgkdaaintngftastldisn- 176  
 QY 115 NNNANIGAKASGNEISAIIGDVLASGASIAIGSDVLYKKTVOQISLPLINGOKAL 174  
 Db 177 enl-karntfted-ekdkaael-vnhgllvvgkdsavnllgkvknevgvlsvng- 232  
 QY 175 NDLYQADTFLQKRYRTHAGHASTRAVGAMSYAKGHFSNAFGRATAECTYSLAVGLAT 234  
 Db 233 llaqklttsdlinpeltysiaapnea-vnlg-dl-fakgnlnvraatlingklsad 289  
 QY 235 AKASSIANGSN-AQAIIGFAATAVGSGTVNLMKRGIALFGSCVLOKNDV-NAANVRX 292  
 Db 290 sveakde-qniylsa-kegeaeigvlsaqnqakgklimtgdkvtlktgavidlsqke 347  
 QY 293 APDDNDPIDRKYKATRNKGAITDFISNGSNGDSIRKLIINWAGASA-DTDAY-NVAOLK 350  
 Db 348 gsetlylgdgergegnqglakktlsleketlnvsgkekygrairavvgdiall-dgn-ina 405  
 QY 351 EAVR-LA-N-ROITFRKGDSDNNRVERKGLKTLTITGAQTS-ALTDHNTGVQNGGLKV 406  
 Db 406 gsgsdiktgyfetsghdlfikdnalvdekewlfdhnvsiaedprlntngindfpt 465  
 QY 407 QLAETLTLK-MV-TT-ENLTANEKTV-GTKPLT-DKGF-TND-MNGIDESKRYLDK 459  
 Db 466 gteasgpkkselktlntltisny--lknawtmllsarkltvnsinlg--sn-shl 520  
 QY 460 DTGHNAGQKITRKLAVGVDDAATYQQLKVNQTAESALQPTFVKVKNGDANDSKX 519  
 Db 521 llnksggrggvgldg-dltsksg-nltysgvv--dv-hkn-fildgflnltasa 574  
 QY 520 ITVGKNNKPDGTQVNLTKLKGNGCVDTTETNGTTFGLNGLVGNSTLNNGLSK 579  
 Db 575 feggnkardaaakivagvtviltgqkdfnanvslngtqglnissvnltnhsq 634  
 QY 580 NTNSNKQIOWGADGITFTTDSNSKPGAGIE-NTTRITRDIGFA-NTGSLDANKRPLTP 637  
 Db 635 t-linsgn-llingctrkntesygwtshdsbnvsaalnetganftfkylissnsglttg 652  
 QY 638 TGINAGKELTNVQSAINPATN-CGOLDFMNRSTANTESKSGSATIKDLYNLSQ-VPLT 695  
 Db 693 yrssagvafvngvngmsfnlkegkvnfkl-kpennm--n-tesplarf--lanitat 745  
 QY 696 FAGDTGPNVTKRIGEL-LKVGKGTATADLTKNNTIGVVAOSTDLSVLAKTISDIDAV 754  
 Db 746 gggsvf-fdiyansgvg-aelkmselnis--n-ganftlnsh-vrgddaefkinkdltin 799  
 QY 755 NTKITLASDKYTVDSGNNTALQNGDLTFSKONGAPATNTSKTIYGVDDKFTDNNGIA 814  
 Db 815 LDGTYITRKD-KVFAKODGSLDKSKPLDKLKVSEVELTITGINAGGAIATGSLNTL 873  
 QY 874 ekaavvlt-eannepngni-rdvyikgsllyvngslstgeaditknltisefk-x 910  
 Db 911 gk-trdlnltgnfngtaeintqgvkl-gnvtdgdlnt-tch-aknqrsilggd 966  
 QY 932 GNATTAKVYDDK-ASKVAVDVNVGTTIHLTGADGNKNOGVATTTLTKADAGDAIN 990

```

Db 967 1-inuksgslntdsndae-ig-iggnisqkgnlttsdsknltkqtlkigigeds 1023
QY 991 FSVNSGDDKALINAKDIALNLTLAGEIRNTKGTADTAQTQVKK-VK-ENGDDNDND 1048
Db 1024 sdatsnaatlktkklkledslsfgfkaeltakdgrdlignsdngsgaeakvfn 1083
QY 1049 TITVGRKADKTNQVNTLKLKGNKGLDQTKKDGVTGVTGTOGLKAGNNTTNNNGLSIK 1108
Db 1084 nrx-dklsiadghvlnskvktssngresnddy-iltakvsnkdltslktv 1141
QY 1109 NTAGNEDQIVGADGVKF-AKVNNGVGAGIDGTRTRITRDEIGFACTNGSLDKSRPHLSMD 1167
Db 1142 nltasek-vttasegslatngkasltktgtslsgntsvtsvatdltksgsklea 1200
QY 1168 GINAGKRTTNQSGEIAONSNDATGKITYDLKTELKISSITAKTQONSHEFSVADE 1227
Db 1201 ksgaanvtaatclgltsgntvntanagdl-tvgngaelnategaatltatgnltt 1259
QY 1228 OGNNFVSNPSSYDTSKTSVDITFAGENGITTKVKNKGVRGIDQITGTLTPKLTVGN 1287
Db 1260 agssit-stkgvdlagngsiasnaavtlttlttltvagsd-ikatsgtlvina 1316
QY 1288 NKGIVTDSQONQNTTGLSLNTLA-NVTNDKGSVRTDQKLIKDEDKTRAASIVDVLSA 1346
Db 1317 k-daklngdasg-dst-e-vnavnasgsgs-vt-aatssvntgldtngt-Inliskd 1369
QY 1347 GFNLOGNEAVDFVSTYTLVNFADGNATTAKTYTDDTSKTSKYVDVAVDDTTLEVDDK 1406
Db 1370 -grntvrltgyekiev-kytqpgvasveevleakvlekykdlse-eretlalgvasvr 1426
QY 1407 LGVKTTLTSTGTGANKFALSNQATGDALVXASDIAVHLNLTSGDIOQAKASQANSAG 1466
Db 1427 fvepn-ncltvn-tqnef 1442
QY 1467 YVDADGNKVITYDSTDNRY 1484

```

RESULT 13  
ID R27745 standard; Protein; 1822 AA.  
AC R27745;  
DT 04-MAR-1993 (first entry)  
DE Extracellular factor related protein.  
KW EF; detection; prevention; screening; diagnostic.  
OS Streptococcus suis type II (non-pathogenic).  
FH Key  
FT peptide  
FT Location/Qualifiers  
FT /note="signal peptide"  
FT /note="mature peptide"  
FT /note="858..861"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="934..937"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="999..1002"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1075..1078"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1264..1267"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1362..1365"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1438..1441"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1514..1517"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1590..1593"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1666..1669"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="1741..1744"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
FT /note="repetitive Asn-Pro-Asn-Leu sequence"  
PD MO9216630-A.  
PD 01-OCT-1992.

```

PF 19-MAR-1992; NL0054.  
PR 21-MAR-1991; NL-000510.  
PA (DIER-) CENT DIERGEENESKONDIG INST.  
PI Smith HE Vecht U;  
DR WPI; 92-349215/42.  
DR N-PSDB; 029471.  
PT Deoxyribonucleic acid encoding virulence characteristic of  
PT Streptococcus suis - useful for antibody and polypeptide for  
PT diagnosing and preventing infections in pigs and humans  
PS Claim 9: Fig 1b; 86pp; English.  
CC The sequence is that of the extracellular factor related protein  
CC from Streptococcus suis type II (non-pathogenic) which allows the  
CC detection and the prevention of infections by S. suis in a more  
CC effective manner than was previously possible. It facilitates  
CC screening of e.g. pigs and elimination of infected and carrier pigs  
CC can then be carried out. The new diagnostic tests can distinguish  
CC between avirulent and virulent strains. It may be used in the prodn.  
CC of a vaccine. See also R27744 and R27746.  
SQ Sequence 1822 AA;  

Query Match 1.8%; Score 240; DB 5; Length 1822;  
Best Local Similarity 18.0%; Pred. No. 6,91e-06;  
Matches 212; Conservative 335; Mismatches 563; Indels 66; Gaps 64;  

Db 658 venpanltpeekaavlaqik-dnad-ne-rlkqjlpdsafvnsdgtvsdys-egvrv 713  
QY 507 VDRKNGDANDSKITIVGKNNKPDGQVNTLKLKGENGVDTETNGVTFGLNONGGLTV 566  
Db 714 dgsdltlkatnlnadtrneakagldtkaekkaekraeafsklddlsaaqrqa 773  
QY 567 -GNS-TLNDGSLSVKNT-NSNKO-IQVG-ADGTPPTDSNSKPGAGIENTRTTRIDGIGF 621  
Db 774 aknavaaaagdalkejdakateakekikattaseindatngeln-lds-aeavgekai 831  
QY 622 ANNTGSLDANKP-R-LTPGINAGK-ELTNQSAINATNGGOLDEPNRISTANTEKSG 678  
Db 832 ngakekelak-aevenkafealekvnnpnllseekyafidikeskavakinaent 890  
QY 679 SAATIKDLVNLNSQVPL-TPAG-DTGPNTYTKRIGETLIKVK-GKKTADDLTKNNIGVADS 735  
Db 891 aestaadeaelaynedvlnaagdalnlkledeekkaaidanpntl-peekakalakv 949  
QY 736 TNSLSLVKIAKTLSPDLAVNTKTLTASDKVYVDSGNTAKIQ-NGLTFKSGNGGAPAT 794  
Db 950 eelvnnaesdlxptetvavedkadkdlakeylgaadg-ak-kyiaenpnltpreek 1007  
QY 795 NSKTIYG-VDGL-KFTDNNGIALDGTTYITRKDVGFAPKQDSLDKSPYLDKDLXGCV 852  
Db 1008 drakkavedavkvtaldakstptevdatsdgvakida-eefkatqk-daknk-ia-x 1063  
QY 853 EITTINGINAGKAITGLSNTLTDATMATGHTVQLGIVSDTKRRASISDVLNAGFNLK 912  
Db 1064 eaesakkaid--dnpnltpdekesanavee-aakvataaidkstpdavqveedgvaa 1120  
QY 913 NNGDAKDFVSTYTLVDFINGMATATKATYTDGKASKAVADVAVDGTTHLHGADGNKQIG 972  
Db 1121 lnltl-akadky--viaakl-adeliklledegaeeakaidasumtneekakialqdv 1176  
QY 973 VKTTTLTKTDKAKGKAINFVNSGDDKALINAKDIALNLTLAGEIRNTKGTADTAQTQ 1032  
Db 1177 -vdkykaeleaarvatneheaittekakaaeagaksj-ttgykeardavakadkel 1234  
QY 1033 QVKVYKENGDDNDADTITVGRKADKTNQVNTLKLKGNKGLDQITKNDGVTTFGINOSGL 1092  
Db 1235 gkealteeaeaklivleakledtrkaiednplnsdedkgaelklldavaklatmrda 1294  
QY 1093 KAGNNTTLNNGSLIKNTWAGNEQIQVGADGVKFAKAVNN-CVYAGAGIDGTRITRDEIGRA 1151  
Db 1295 dkr-tgeakagadaleketqkiad-kaa-idrtlt-lvkdgealeatkqaktkiak 1350  
QY 1152 GTNSGLDKSRPH-LSKGINAGGKIKITINIOSGEIAONSNDATVAGKITYDLKTELKISS 1210  
Db 1351 daaakeaiaenpnltdaekftt--davda-evakandaiaatspadvqkeedagvaa 1407

```

Query Match 1.7% Score 229, DB 8, Length 1477;  
Best Local Similarity 19.0% Pred. No. 3,02e-05;  
Matches 153; Conservative 229; Mismatches 374; Indels 48; Gaps 46;

Db 584 daanakivagvtvri-tgegekdfannvsi-ngtqkylisvnmthlslgslinsgn 641  
QY 1255 ENGTTTIVNKGVRVGDITKGLTTPKLTIVGNNGKGI-VIDSQNG-QNTITLSLTLN 1312  
Db 642 ftingtrkntsywqtdshwvsalnlet-ganftilyisnsgkyltgyrssaayn 700  
QY 1313 VTNDKSVRTTEOGKRIKDEKKTAAISIVVLSAGFN-LQ-GNGEAVDFSTVDTVFAD 1370

DR R41728 standard; Protein; 1477 AA.  
AC R41728: 26-APR-1994 (first entry)  
DE High molecular weight protein 2 (HMW2).  
KW HMW, high molecular weight protein; virus; vaccine; influenza;  
KW epitope; immunity; haemophilus influenzae; gene cluster.  
OS Haemophilus influenzae.  
PN W09J19090-2.  
PD 30-SEP-1993.  
PF 16-MAR-1993; U02166.  
PR 16-MAR-1992; GB-005704.  
PT (BARE/) BARENKAMP S J.  
PI (INRM) INSEEM INST NAT SANTE & RECH MEDICALE.  
PI Barenkamp SJ.  
PI WPI: 93-320683/40.  
DR N-PSDB; Q49509.  
PT High molecular weight surface proteins - of non-typeable  
PS Claim 4: Figure 4; 100pp: English.  
CC The isolation and purification of the high molecular weight protein  
CC enables the identification of the major protective epitopes of the  
CC protein by conventional epitope mapping. These epitopes can then be  
CC synthesised using standard techniques and incorporated into fully  
CC synthetic or recombinant vaccines. This sequence is claimed to be  
CC the same as that given in R41724 (High molecular weight protein 2)  
CC although it does differ slightly.  
SQ Sequence 1477 AA.

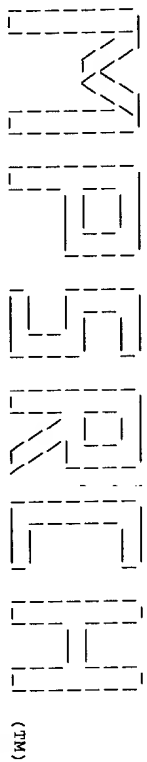
Query Match 1.7% Score 229, DB 8, Length 1477;  
Best Local Similarity 19.0% Pred. No. 3,02e-05;  
Matches 153; Conservative 229; Mismatches 374; Indels 48; Gaps 46;

Db 584 daanakivagvtvri-tgegekdfannvsi-ngtqkylisvnmthlslgslinsgn 641  
QY 1255 ENGTTTIVNKGVRVGDITKGLTTPKLTIVGNNGKGI-VIDSQNG-QNTITLSLTLN 1312  
Db 642 ftingtrkntsywqtdshwvsalnlet-ganftilyisnsgkyltgyrssaayn 700  
QY 1313 VTNDKSVRTTEOGKRIKDEKKTAAISIVVLSAGFN-LQ-GNGEAVDFSTVDTVFAD 1370

Db 701 fngvngmsfn-lkegavnkflkxnen-mnt-ekpjlifrlanltatggsvfiddyan 757  
QY 1371 GNATTAKVYDDTSKTSKVVYDVNDTTEVKKKLGK-TTTLTSTGTAANKFAL-SN 1428  
Db 758 hsgrgaelmseinsnganftlnshvr-gddafinkdl-tlnatsnflrqtckdly 815  
QY 1429 QATGALVYASDI-VAH-LN-TLSGDIOTAKGASQANSAGVADCKVITYDSIDNKY 1485  
Db 816 dgyarnalnstynlsl--lgnv-ltggqnsstlgnltlekaanvleannapngni 872  
QY 1486 QAKNDGTVDKTEVAKDKLVAAQOTPDGTALQMNKSVINKEQVNDANKKOGINEDNAFV 1545  
Db 873 rdtvklgslvngslst-ge-nadikgnltis-essafkxtrdtlnlgnfngt 928  
QY 1546 KG-LEKAASDNKTKNAVVGDLNVAQOTPLTFADDTGTTAKKLGELTLIKGGQDTNKL 1604  
Db 929 aenltgqvklg-nytnqgd-lilt-haknrgtsllgdltnkkslnltsmdne- 984  
QY 1605 TDNNIGVVGTDGFTVFLAKDLTLNLSVNAAGCTKIDDKGVSPVDSGQAKANTPVLASNG 1664  
Db 985 lqignisqkegnltsdsklnltqitlkgldgedssdsanltlktelkted 1044  
QY 1665 IDLGKRVISNVGKGTDTDAANV-QQLNEVRNLDLGNRGNNDNAGN-QVNIADIK-KDP 1721  
Db 1045 ltsifnkaeltakgdrdltnsgndgnsqaeakvtfnnvkslsadqhnvlnskv 1104  
QY 1722 NSGSSSNRT-VT-KAGTVLG-GKGNNDTEKLATGQVGVDDGNANGDLSVWVKTO-K 1777  
Db 1105 tssnggrees-nadnltg-ltlt-aknve-vnkdlstltnltasekvttaagstlnat 1160  
QY 1778 DGSKKALLATYRNAQOTNYLTNPAEAIIDRINEOGIRFHVNDGNEPVGRRNGIDSSA 1837  
Db 1161 nykasltk-tgdisgtisngtvsatgdltksgskleasgeanvsacgtl-gftl 1218  
QY 1838 SGRSVAIIGFOAKADGAAVAIGROTAGNOSTAIDNQA-TGDDSTALIGCNVYAGKH 1896  
Db 1219 sgntvn-vtanegd-ltvgnaelinategaatltaagntlteagssltstkyqvdllq 1276  
QY 1897 SGALGDSITYKADNSYSGVNNNOFTDQIDVGVGNNTTYESNVALGSALISAGTH 1956  
Db 1277 ngsiagsinaanvltntgtltvvaagdkatqglvinakda-klndaagdsdevna 1334  
QY 1957 AGTOAKKSDGTAGTTTATGATGTVKGFAGOTAVGASVGSAGERRIIONVAGEVSATST 2016  
Db 1335 vndvfgsvtaatsssvnyngdin 1358  
QY 2017 DAVNG-SQLYKATOGIANATNEID 2039

RESULT 15  
ID W30293 standard; Protein; 1536 AA.  
AC W30293:  
DE 14-APR-1998 (first entry)  
DE Non-typeable Haemophilus high mol.wt. surface protein HMW1.  
KW Non-typeable Haemophilus; high molecular weight surface protein;  
KW HMW1; hmw1a gene; immunogen; vaccine; otitis media.  
OS Haemophilus influenzae strain 12.  
FH Key Location/Qualifiers  
FT Misc\_difference 4  
FT /note= "encoded by CTA"  
FT /note= "encoded by GAT"  
FT /note= "encoded by AAG"  
FT Misc\_difference 363  
FT /note= "encoded by AAG"  
PI W09J36914-A1.  
PD 09-OCT-1997.  
PF 01-APR-1997; U04707.  
PR 01-APR-1996; US-617697.  
PT (BARE/) BARENKAMP S J.  
PI Barenkamp SJ.  
PI WPI: 97-503038/46.  
DR N-PSDB; T90994 and T90996.  
DR High molecular weight proteins of non-typeable Haemophilus





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Oct 28 11:21:24 1998; MasPar time 98.78 Seconds  
785.129 Million cell updates/sec

Tabular output not generated.

Title: >US-08-968-685-10  
Description: (1-2123) from US08968685.pep  
Perfect Score: 13368  
Sequence: 1 MNHRYKVFNKATGTFFMAVA.....NGSADTQGHVGAAGAFHF 2123

Scoring table:  
Gap 11  
PAM 150

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r56  
1:p1r1 2:p1r2 3:p1r3 4:p1r4 5:nr13d

Statistics: Mean 56.332; Variance 187.426; scale 0.301

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	339	2.5	2249	2 A41477	190K surface antigen	4.56e-21
2	272	2.0	1109	2 A56143	surface-array protein	7.95e-14
3	251	1.9	1822	2 S33441	ER protein - streptoc	1.24e-11
4	234	1.8	920	2 I40614	surface array protein	6.86e-10
5	243	1.8	1176	2 A33856	surface-layer 125k pr	8.28e-11
6	234	1.8	1477	2 B43855	high-molecular-weight	6.86e-10
7	221	1.7	1577	2 A35140	hemolysin A precursor	1.40e-08
8	209	1.6	455	2 S04912	yopA protein - Yersin	2.15e-07
9	217	1.6	555	2 S70791	yopA protein - Mycop	3.49e-08
10	210	1.6	936	2 I40711	sapB protein - Campy	1.72e-07
11	217	1.6	1536	2 A43855	high-molecular-weight	3.49e-08
12	206	1.5	422	2 S04911	yopA protein - Yersin	4.23e-07
13	207	1.5	434	2 S04534	invasin precursor - Y	3.38e-07
14	207	1.5	713	1 UMMS	period clock protein	3.38e-07
15	204	1.5	1026	2 A48995	paracrySTALLINE surfa	6.62e-07
16	204	1.5	3029	2 S76109	hypothetical protein	1.18e-05
17	191	1.4	298	2 I64138	adhesin homolog H1173	8.33e-06
18	192	1.4	485	2 A60610	circumsporozoite prot	9.48e-06
19	192	1.4	852	2 D64891	hypothetical protein	7.61e-06
20	193	1.4	1150	2 S66852	hypothetical protein	3.51e-05
21	186	1.4	1150	2 A40009	mucln, submaxillary -	1.18e-05
22	191	1.4	1175	2 S70843	hemolytic protein hnd	4.59e-04
23	174	1.3	351	2 U00056	flagellin - Serratia	

ENTRY	TITLE	ORGANISM	DATE	RESULT
24	180	1.3	429	2 A54504
25	168	1.3	571	1 UMFR
26	175	1.3	670	2 F36791
27	175	1.3	792	2 S70305
28	177	1.3	995	2 S50358
29	170	1.3	1302	2 JC5009
30	168	1.3	1965	2 S75200
31	161	1.2	398	1 OZ20AS
32	160	1.2	401	1 QXBP1L
33	162	1.2	489	2 JC5755
34	162	1.2	518	2 A55864
35	161	1.2	573	2 A42474
36	161	1.2	573	2 A35146
37	164	1.2	595	2 B48558
38	163	1.2	595	2 A48658
39	164	1.2	835	2 JC6140
40	165	1.2	943	2 S31132
41	161	1.2	1322	2 S07053
42	163	1.2	1569	2 A65044
43	160	1.2	1612	2 A36473
44	166	1.2	1645	2 JN0896
45	162	1.2	2383	2 D64962

ALIGNMENTS

RESULT 1  
ENTRY A41477 #type complete  
TITLE 190K surface antigen - Rickettsia rickettsii  
ORGANISM #formal name Rickettsia rickettsii  
DATE 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 09-Sep-1997

ACCESSIONS A41477  
REFERENCE A41477  
#authors Anderson, B.E.; McDonald, G.A.; Jones, D.C.; Regnery, R.L.  
#journal Infect. Immun. (1990) 58:2760-2769  
#title A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences.  
#cross-references MUID:90354033  
#accession A41477  
#molecule\_type DNA

##residues 1-2249 #label AND  
##cross-references GB:M31227; NID:9152465; PID:9152466  
#note the authors translated the codon GAT for residue 430 as Gly, and CAA for residue 501 as Asn  
surface antigen; tandem repeat

KEYWORDS  
FEATURE  
1-20

SUMMARY  
#domain signal sequence (uncleaved) #status predicted  
#length 2249 #molecular-weight 224332 #checksum 9532

Query Match 2.5%; Score 339; DB 2; Length 2249;  
Best Local Similarity 21.0%; Pred. No. 4.56e-21;  
Matches 221; Conservative 251; Mismatches 522; Indels 58; Gaps 53;

Db	171	LCGANAALIOSAAPS-KITLAGNIDGGITVTK-DAINGTGNINATLVVAGT- 227
QY	724	LRKNMIGVADSTDSLVKLAKTLSDDAVNTKTLNSDVTYVDSGNMNAKLONGDITF 783
Db	228	ATLGGAVIKATTTKLTNNAVLTITNANAVALTGAIDNTGGDNVNLNGALSQVTGDI 287
QY	784	SKONTGAPATPKSKITTYGVGDKFTDNGIALDGTYYTK-DKVGAFKQDSLDKSPYL 842
Db	288	GNTN-SLATISVAGATFATLGAIVKATTTKLTDAASAVK-FTNPVYVGTALDNTGNANN 345
QY	843	DKDKLKVGEVEITTINGINAGKATITGLSNTLTDAATNATTHVTOGLIVDSYDKTRASIG 902
Db	346	IYTFPG-NSTVIGNGN-TNLATVN-VGAGLQVQGV-KKANTINTLNASAVTFETNP 401
QY	903	DVLNAGFNKKNNDKADKPYSTIDYDFINGNNTAKYVYDGRASVAVADVDDTTI-HL 961
Db	402	VVVTGAIIDNTGNANNIV-T-FTGNSYVT-G-DIGNTALATV-NVAGATATLGAV--I 454

```

QY 962 TGAGCNKNOICVKTTLTKTDKADKAINFSVNSGDDKALINAKDINLTAGEIRNT 1021
Db 455 KATP-TKL-T-NAASVL-TLTNANAVLTGAIIDNTGGDNVGLNLNG--ALS-OVT--GN 505
QY 1022 KGTADTALOTFOVKKVENGGDDNDADITTYGKADAKTNOVWTLKCKKNGLDIDQTKNDGT 1081
Db 506 IG-NTNSLATSISVAGTA-TLGGAVIKATTTK-LTD-AASAKFT--NPVVTGAIDNTG 559
QY 1082 VTGIFNOSGKAGNNNTLLNNGSLISIKNTAGNEQIQVAGGVFAKVNNGSVAGIDGT 1141
Db 560 NANGIATFTGNTSTVTDIGNTSLATISVAGATLGGATLGGAVIKATTKLTNNAVLTITN 619
QY 1142 RITDELGFACNTG-SLDKSRPH-LSKDGINAGCKITTIQSGEIAQNSMDVATGKITYD 1199
Db 620 ANAVLTGAIIDNTGGDNVGLNLNGALSOYTGIDGNTSLATISVAGATLGGAVIKAT 679
QY 1200 LKTELEKKSISTAKTQNSLHEFSVADEQ--GNFTVSNPYSYDTSKTSQVITFAENG 1258
Db 680 TTKTNNAVSAVKFTNPVVTGAIIDNTGNAN-NGIYV-F-TGNTSTVTDIGNTNALTAVN 736
QY 1259 TTKVKNKGVVRGIDQTKGLTTPKLTGWNNGKGVIDSQNGQNTITG-LSNTLANVTNDK 1317
Db 737 GAGTATLGGAVIKATTKLTNNAVLTGAIIDNTGGDNVGLNLNGALSOY 796
QY 1318 GSVRTFOGKIIDED-K-TRASIVDVLSAGFNLOGEAVDFVSTYDVFADG-NAT 1374
Db 797 TGDIGNTSLATISVAGATLGGAVIKATTKLTNNAVLTGAIIDNTGG 856
QY 1375 TAVYTDITDTSKTSKVYDVN-VDDTTEVKKKLGKVTLLTSTGANKFALSNQATGD 1433
Db 857 DNVGLNLNGALSOYTGIDGNTSLATISVAGATLGGAVI-KATTKLTNNAVLT 914
QY 1434 ALVKASIVAHMLTSLGDIOTAKGASQANSAGYDADGKVTYDSTDKKYQAKKDGTV 1493
Db 915 TNANAVLTGAIIDNTGGDNVGLNLNGALSOYTGIDGNTSLATISVAGATLGGAVIK 974
QY 1494 DKTEVAKDKLVAAQTPD-GTLQNNVKSIVIKKEQVNDAN-KKQINDNDNFVGLERA 1551
Db 975 ATTTKLTDASAAYKFTNPVVTGAIIDNTGNANGIYTFGNTSTVGNVGNATNALAVNG 1034
QY 1552 ASQNKTKNAVTVGLDNLAAQCT-PLTFAGDGTGTAKKLEETLLIKGGQDTNKLTDNNG 1610
Db 1035 AGLLQVGGVVK-ANT-I-MLTNDAVFTNPVVTGAIIDNTGNANGIYTFGNTS--TNT 1091
QY 1611 V-VAGDGFVTLAKDLTNLNSVNAAGTKIDDKGS-FVDSGQAKANTPVLSANGLDIG 1668
Db 1092 GNGNTNALATVNVAGLLQVGGVVKANTINLTNDAVFTNPVVTGAIIDNTGNAN 1151
QY 1669 GKVISNVGKGTDTDAANYO-QLNEVYR-MLTGLNAGNDNADGNVNIADIKDPNSGSS 1726
Db 1152 GIYFTGNTSTVTDIGNTNA--LATVNVGAGI 1181
QY 1727 SNRTVIKAGTVLGGKGNNDTEKLTATGVOGV 1758

```

```

RESULT 2
ENTRY #type complete
TITLE surface-array protein homolog sap2 - Campylobacter fetus
ORGANISM #formal_name Campylobacter fetus
DATE 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change
09-Sep-1997
ACCESSIONS
REFERENCE A56143
#authors Dworkin, J.; Tummuru, M.K.R.; Blaser, M.J.
#journal J. Bacteriol. (1995) 177:1734-1741
#title A lipopolysaccharide-binding domain of the Campylobacter
fetus S-layer protein resides within the conserved N
terminus of a family of silent and divergent homologs.
#accession A56143
#status preliminary
#molecule_type DNA
#residues 1-1109 ##label DMO

```

```

#cross-references GB:S76860; NID:9913763; PTD:9913764
#note this gene appeared to be silent in strain 82-40 LP3
GENETICS
#gene sap2
SUMMARY #length 1109 #molecular-weight 11805 #checksum 3205
Query Match 2.0% Score 272; DB 2: Length 1109;
Best Local Similarity 20.5%; Pred. No. 7.95e-14;
Matches 195; Conservative 25; Mismatches 434; Indels 69; Gaps 60;
Db 75 ALGNTSDVDGINVTKATITGGEEFTDSKGVNIVAS--LSKDLIGAMINSMVNGSAES 132
QY 240 SIYVSNNAQAIQFATAVGSGTQVNLNGIALGFGSOVLQNDNVNA-ANVRAA-PDN 297
Db 133 KAIFPAK-AASDVPADLTICKDISGDEGTSKLISEINASDLPVKSEIDGKESID 191
QY 298 OPT-DNRKATFKKGAATDFVSIQNSGNDSTRKRIINVGASADTDAYN--VQOLKEAVR 354
Db 192 BAGLNKIALTTENDTITGTEG-GDLISGVGTAAESTLNFQDKIDGAGNDVLRVDKNN 250
QY 355 LAN-RQIFKGDSDSNRREKELGKTLT-ITGGAQTSALT-DHNGVQVQNDGLKVOIAET 411
Db 251 FKGLK--DDGTTKINEKISTLNSSVSNRTPD-AKIDGL-QIYA-LSGEKISVTMLANI 305
QY 412 LTLKMTTENTLANKETVGTGKTLTDKIGFTNDMNGIDESPYPIDKGTGTHAGG-QKI 470
Db 306 VDVEVNGFKGTNFVDSIYADKVLDS-ADYQNLKVGVAGKASVAITDKIETILNLT 364
QY 471 TKL-TAGVVDDAATYQOL-KKVQNTAESALQTFYAKVKDKNSD-A-NSDKITTYGKN 526
Db 365 TGSGSFVSADVASISVKGANLISATGAK-TTLTDASSFEGALADLST-SASVTSIKGG 422
QY 527 K-PDG---TQVNTLKLKENGVDVTEETNGTVFGLNQNG-LTVGNSLTINNGLSVKNT 581
Db 423 NGNKKITIKVAVVVALDGCAGNDELVIKSGTADTLOPTLTINEKYTIDGNTKDLTSLK 482
QY 582 NSNKOIYVAGDGIET-DISNKRKAGIEITRTITRIGIFANNTSLDANKRPLPTGI 640
Db 483 KAOVTESEFKINAKVYTESNGVETVNIILANNATK---AVINDE-SIKTINFSQVD 538
QY 641 NAGG-KELTNVQASLNPATNGOLDPEFNRLSTANTKSSGAATIKLVLNSQVPLTFACD 699
Db 539 KGASVAAK-GKIYADKATBELTINSKVTLASDAVQAAATKIDINAADTVGLTGVYA 597
QY 700 TGPVNTKKEIILKVGKGK-TT-ADDLTKNNIGVADSTDSNLTVLAK-TLS-DLDAVN 755
Db 598 KLTLDTLVNKGAFALGANATDLDVSK-NLSVNTBEAFSIAATATSIKNNINSL---NGV 653
QY 756 TKT-LTASDKVYVD-SCNNATKALQNDLTFSSKONTGATPATNSKTIYGVGDKLFTDNNGI 813
Db 654 SADLNS-VNVTATLASLEANINVSSEFKLGTITAK-GVDENIENVAGLTGATTSSTG 711
QY 814 ALDGTYYITKDKYFAKQDSDLSKSPY-LDKDKLKVGEVELTITNGINNKG-KALTGLSN 871
Db 712 NASYIISATGANT-LGANSATQGNLTLAGNTLGNITIGALKGDIYSVDLGGVLTINS 770
QY 872 TLTDAINATGHVTOGLIVDSIDTKTRAASIGDVL-NAGFN-LKNNGDADKFSTYDVF 929
Db 771 DANNKVISITSEVYTGSEISKVVEITTAAGCTDINAQVIGAAADDLTIIGGDQOT 830
QY 930 -INGNAT-TAK-VTYGKA-SKAYAVVNDGTTHLTGADGKKNQIGVKTTLITLTDKGG 985
Db 831 ITASG-DLSSGTLTLTLEATKLS-SLD-ISG-VKGITGNVAIELGKAVQNGKTDVVOG 886
QY 986 DKAINFSVNSGD-DKALINAKDIADLNLAGEIRNTKTADTAL-QTFQVKKVENGGD 1043
Db 887 SDAEOIYTTASALDIDISGDLGAGATTVTPTAAADKTIDL-SGLSATGTLAS 945
QY 1044 DNDADITIVGKDKTINQV-TLKL-KGKNGLDIQTKNGGTTFGINTQSGKAGNNITLN 1101
Db 946 TITLVANATATSVKSLGADITTVASANKAV-AIDLKGPDAIDVDVS 993
QY 1102 NNGLSIKNTA-GNEQIQVAGADGVKFAKVNNGVAGAGIDGTRITRIDEIG 1149

```

RESULT	3	
ENTRY	S33441	#type complete
TITLE	EF protein - Streptococcus suis	
ORGANISM	06-Jan-1995	#formal name Streptococcus suis
DATE	09-Sep-1997	#sequence revision 06-Jan-1995 #text_change
ACCSSIONS	S33441	
AUTHORS	S33441	
#authors	Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.	
#submission	Submitted to the EMBL Data Library, May 1993	
#description	Repeats in an extracellular protein of wek-pathogenic strains are absent in pathogenic strains.	
#accession	S33441	
#status	Preliminary	
#molecule_type	DNA	
#residues	1-1822	#label SMI
#cross-references	EMBL:X71880, NID:9298031, PUD:9298032	
SUMMARY	#length 1822 #molecular weight 192031 #checksum 8006	
Query Match	1.9%; Score 251; DB 21; Length 1822;	
Best Local Similarity	18.1%; Pred. No. 1.24e-11;	
Matches	213; Conservative 340; Mismatches 557; Indels 68; Gaps 64;	
Db	568	VENPANTLPEKAAVIAIOIK-DNAD-NE-RLKGLPDSATVSDGTVSDYS-AGGVAV 713
Qy	507	VDKNNDNDNSDKITTVGKNNRPDGTQVNTLKLKNGECVDYTTNGTVTFGLQNGLTV 566
Db	714	DCATDIIKNAATNLADTFENOKAEIDFKLAEKKAIEKRDFAFSKIDDISLRARORA 773
Qy	567	-GNS-TLWNGDSYKNT-NSNK-QIQVG-ADGITFTDISNKKGAGACENTRTTRDGI 621
Db	774	AKDAVAAAAAGALKELDKATEAKEDIKATVTASEINDAKTNGEIN-LDS-AEAVEKAI 831
Qy	622	ANNTGSLDANKP-R-LPTGTGINAGGK-ELTNVQSAINPATNGGLDFMRLSTANTEKSG 678
Db	833	NOAKKEKELAK-AEVENKAFELAEKYNPNPILLEEKAYVDIIEKSEKVEVAKINNAEMT 890
Qy	679	SAATTKDLYNSQVPL-TFAG-DTGPNYTKRLGILYVK-GGKTTADDTLKKNNIGVADS 735
Db	891	AEITTAIDEAEIAYVEDYVINAADALDKLEKDESEETKAAIDANPNLT-PEEKAKIAYV 949
Qy	736	TNLSLTVLKIAKLSLDLADVNTKTLTASQKAYVDSGNNPATLQ-NGDLTFSKONTGATPAT 794
Db	950	EELVNNAESDLSKPTPTVOAVEDKADKDIKAYELQAAADG-AK-KGTANPNLPEEK 1007
Qy	795	NSKTIYG-VDGL-KFTDNNGIALDGTITITKDKGFKAQODSLSKSPYLDKRLKVGEE 852
Db	1008	DVAKKAVEADVAVDAIDAKASTPEVEVATSDGKAKIDA-EPEKAKQK-DANKK-IA-K 1063
Qy	853	EITTNINNGKRAITGLSNTLTDAATNATGGHTAYTOLGIYDSIDTKRAASIGDVLNAGNLIK 912
Db	1064	EAEASAKAID--DNNPLTPDEKESAKNAVEE-AAKVATAIDAKASTPDVAQVEEDGVAA 1120
Qy	913	NNGDADAEVSTIDYDFEINGNATTAKYTVYDGKASKAVAYADVNDGTIHLTGADGNKNQIG 972
Db	1121	INLIT-AKADAG--VIAAKL-ADIRKLEDKQAEEAKAIDASTMTBEAKILAKKALADV 1176
Qy	973	VKTTTLTJTKKADGKRAINFSVNSGGDKLLINKADIDADLNTLAGEIRNTGATATLQTF 1032
Db	1177	-VYKGAETEDAARVATNEIHEATTEKAKAAELGKSL-TDVGKEARDAVELANDKEL 1234
Qy	1033	QYKRVENGGDDNDADITVIGKDATNTQVNTLKLKGNKGLDIQNNKQGVTFPGINTSG 1092
Db	1235	AKFAIRTEEBEATKIVKLEADTRKAIEDNPULSDCKOAEIKKLTDAVAK-TLATIRDN 1293
Qy	1093	KAGNNPTLLNNGISIKNTAGNEQIOVGADGVAFAYVNN-GVVGAGIDGTITITITDEIGFA 1151
Db	1294	ADKRTOEAKAQLADLEAKAFKQKAD-KAA-IDRLTI-LYKQGELEATRODAKNTIAK 1350
Qy	1152	GINGSIDKSKPH-LSDQGINAGGKTIINQSEELIQNNSNDNAVTEGKITDLKTELEKNISS 1210

Db	1351	DAKAKTAISNNPLTDAEKKFTT--DAVDA-EVAKANLISAAISPADVQKEEDAGYAA	1407
Qy	1211	TAKTAQNSL-HEFSVDEQSNNTFVSNPSSSDTSTSDVITFA-GENGITTAKVGVR	1268
Db	1408	IAEVLPLAAKQDAKNNJAKDPAAAKKAISNPLTDAEKKFTTDAVDAEVAKANDAIISA	1467
Qy	1269	VG--I-DQTKGLTLPKLTGVNNGNGKGIVIDSQN-GQNTITGLSNLT-ANVTNDKGSVRT	1322
Db	1468	TSPADVOKEEDAGYAAIAEVLPLAAKQDAKNNJAKSDDAKSAIDANPNLTDAEKSARK	1527
Qy	1323	TEOGKIIDECKRTBAIVLVSFG-FNLQGN-GEAVDPV-STYDTPV-NEAOGNATTAKV	1378
Db	1528	AVDADAKATDAIDASTSPVEQASAEEDKGVGSIADOVPLAAKQDAKNNJAKEVAA-A-KE	1585
Qy	1379	TYDDTSKTSKRYVYVNDOTTIE-VNDKRLGKVTTLTISTGIGANKFALSNQATGALVK	1437
Db	1586	AIDANPNLSDAKEASAKKAVDADAKATTAIDASTSPVEQASAEEDKGVGSIADOVPLAAK	1645
Qy	1438	ASDIYAHLNTLSGDIQTKAGKQANSAGVADGKMKYIYDSDNNKYQAKNDGTVDKAK	1497
Db	1646	ODAKKKIAKESDAKSAIDANPNLTDAEKSARKAVDADAKATTAIDASTSPVEQASAE	1705
Qy	1498	EVAKDKLVAQOQTDGTL-QOMNKSVYNN-EO-VNDANKKQGINEDNAFVKGLEKAAAD	1554
Db	1706	DKG-VGAIKAPLIDPLAAKQDAKNNJAKKAESAKSYIDSNPMLTDAKAEAK-SEIDKAVEE	1763
Qy	1555	NKTKMAATVADLNAVQPTLT-FAGDTGTAKKLGTELTIKGQGDPTNKLTDNNITGVVA	1613
Db	1764	ATVLINGRYTQELBKIKLPMALIRKAAVTPVDPN 1801	
Qy	1614	GTDGFT-VKLAKDLTNLNSVAGGTIKIDDKGVSEFVDS 1650	
RESULT	4		
ENTRY	140614	#type complete	
TITLE		surface array protein - Campylobacter fetus	
ORGANISM		#formal_name Campylobacter fetus	
DATE		#cross-references revision 12-Aug-1996 #text-change 12-Aug-1996	
ACCESSIONS	140614		
REFERENCE	140614		
#authors		Tumkur, M.K.R.; Blaser, M.J.	
#journal		Proc. Natl. Acad. Sci. U.S.A. (1993) 90:7265-7269	
#title		Rearrangement of sapa homologs with conserved and variable regions in Campylobacter fetus.	
#cross-references		MUID:93348254	
#accession	140614		
#status		preliminary; translated from GB/EMBL/DBJ	
#molecule_type	DNA		
#residues	1-920	#label RES	
#cross-references		GB:LI5800; NID:9289533; PID:9289534	
GENETICS			
#gene	sapa		
SUMMARY		#length 920 #molecular-weight 94996 #checksum 2434	
Query Match		1.8%: Score 234; DB 2; Length 920;	
Best Local Similarity		20.7%: Pred. No. 6,8e-10;	
Matches	151;	Conservative 201; Mismatches 323; Indels 56; Gaps 54;	
Db	62	GNNKDPVTIYSIALGNSTDVVG-INYTKAITGGGEFTSGKGNVISAASL-SKGDLLGA 119	
Qy	915	GDKKDVSTYDTPVPLNGNATTAKTATYGGKSKAYID-VNVDGTTIHLTGADGKNKNGIV 973	
Db	120	MINSNNGSAAESKAI-PEAKAASDYADAT-LGRDISGLD-EGTSSKLISEINSAS-D 175	
Qy	974	KTTTLTKTDKAGKAIINFSVNGDDKALINAKDIADINLTLAGIIRNKGTADTALOTFQ 1033	
Db	176	LDRVASE-IDALKSELPNGSGTYDLEGN-MLKQTD-LD-DEF-NGTVYLGGINSTL 230	
Qy	1034	VKRVKENGGDDDDADITTVGKDAKTQNVATLKLKGNGLDIQTKNGDITVTFGINTQ-SGL 1092	
Db	231	TARDPVDAAGANDTINIVYGLSNSNV-TD-LKISDLSALLGVTNVEKLNIISEVKAAD 288	



QY 1093 KAGNNVTLNNGLSTIKRNFAG-NEQIOVADGVKFAKVNNGVAGAGIDGTTTRTRDELGEFA 1151  
Db 289 GGGLTINQYKS-VSLN-I-VGETKIADTADATKIDKASGNVTVFAEAVK-DLSINRNS 344  
QY 1152 GTNGSLDKSKPHLSMDGINAGCKKITINOSGSIANSNDANTGGITDLKELE-NKISS 1210  
Db 345 EYTIANAATKALENTITINAKSLTATNFDG-DILKSVTLNVLGDTNACAAFDKGS 403  
QY 1211 TAKTQNSLHEF-SVADBOGNFTVSNPSSYDTSKIDSVITLFA-GENGITTKYV-KGVV 1267  
Db 404 TLNPNVNSAOTPAASKIAHITSSAETLNLKALSGTATVAITVTSVVAIINNA-DVN 462  
QY 1268 RRGIDQT-KGLTTP-KLVGNNNGKGIYIDSONGONTTGL-SYTLAVIWDKGSVPTTE 1324  
Db 463 AFLITKGGDMNPGHADLQNDSEFTTIFVKGNGKELTLNAGDLVDKIDITGFGSNAKVS 522  
QY 1325 QGKIIRKDE-DKTRA-ASIVDVLSAGFNLOGGEAVDF-VSTYDYV-NF-ADGNMTAKVT 1379  
Db 523 FGDDTSADGSQLSVKTGAGNDTINLEAKL--KAGSLIDGEGDITIMKASALADATL 580  
QY 1380 Y-D-DTSKTSKVYDVNDVDTTIEVKDKILGVKTTTLTGTGANKKALSQATGDA--L 1435  
Db 581 GMIKNI-ENV-TVSPDALSANTDVASASFPVNTGL-ADKTDAPFELTVNKN-QTIDIOSKE 636  
QY 1436 VKASDIYVHMLTSGDIDTAKGASQAN-SSAGYVADAGCKKIYISTDKKIYQAKNDGTVD 1494  
Db 637 MAKSOILITKNDMSGSDIVYIVLMAKAITNOGDKRVAQAOTKGLKIDGIESVNTS 636  
QY 1495 KTK-EVANDKLVQAQEPDGTGLAOMNVKSYNK-EQ-VNDANKKQINEDNAFYK-GLEK 1550  
Db 697 VAKDNTANTLWIDPSSGCTGANKIYISGDDITVAASVATGKIKDLDSALTKL 756  
QY 1551 AASDNKKTNNA-V-TVGD-LNAVAQTPUTRAGDTGTAKKILGETL-TIKG-QGTD-TNKL 1604  
Db 757 TEDASVKNLAS 767  
QY 1605 T-DNNITGVAG 1614

RESULT 5  
ENTRY A33856 #type complete  
TITLE surface-layer 125K protein precursor - Bacillus sphaericus  
ORGANISM #formal\_name Bacillus sphaericus  
DATE 09-Mar-1990 #sequence\_revision 09-Mar-1990 #text\_change 09-Sep-1997  
ACCESSIONS A33856  
REFERENCE A33856  
#authors Bowditch, R.D.; Baumann, P.; Yousten, A.A.  
#journal J. Bacteriol. (1989) 171:4178-4188  
#title Cloning and sequencing of the gene encoding a 125-kilodalton surface-layer protein from Bacillus sphaericus 2362 and of a related cryptic gene.  
#cross-references MIMD:89327128  
#accession A33856  
#status Preliminary  
#molecule\_type DNA  
#residues 1-1176 #label BCW  
#cross-references GB:M28361; NID:g341911; PID:g956012  
SUMMARY #length 1176 #molecular\_weight 125225 #checksum 3319

Query Match 1.8%; Score 243; DB 2; Length 1176;  
Best Local Similarity 20.7%; Pred. No. 8.28e-11;  
Matches 186; Conservative 228; Mismatches 426; Indels 57; Gaps 55;

Db 268 VYTTATVKSVAATNKEVYVDFGVKETAADAAVNAIKSGTKTSVSLAANDKTATV 327  
QY 231 LRTATKAAASSIAGS-NAQAIGFAATAVAGSTGVNLNRGIALGFCGQVLOKDNVMAAV 289  
Db 328 TLADKLNNKADAISISNVAGDKEI-NYKNVEFT-AVDNKEIPEYEVKSLGKAKATV- 384  
QY 290 RAAPADNDQIDNRYKATFNGATIDVFSIGNSGNDSTIRKTIIVGASGA-DTDAVNAQ 348

Db 385 LSEPVENSTNETTLIDKAYFGNVVWAGNKTIVILTPYSSALSVDKHLTVSGAKDFAG 444  
QY 349 LKEAV-RANRQITTFEGDSSNNRVEKGLG-KTLITIG-GAOTSLDTHHIGVYONGDLK 405  
Db 445 FVSLNST-HEFKVY-ED-KEAPYTEATATLETVTLFSESDIMDTYKASNVYKSGDS 500  
QY 406 -VOLAEITLSLKAVTENTLANEKVAVGKTRLTLDKIGFT-N-DNNGIDESKPYDK-DT 461  
Db 501 KKEASEERADKKYFVFKGSEKTLPTGKDVYVED-IKQSDKIKAKDKIVYTPPID 559  
QY 462 GHAGG-QKITKLTAGVDDDAITGOLKKNVQTAESALQITFYKVKVKNGDADSKII 520  
Db 560 QPREPVKATYALDEKTIKVFESKTYDGESEAIK-TGNTYTKDKDQKVSVDKTVDSKSK 618  
QY 521 -FVGNKKNKPDGTQVNTLKLKGENGVDTTENTGTYTFLGNQNGTLVNSLNDGLSVK 579  
Db 619 SVIIDLKSVSGENTIT-IRKVKATATKNTM-LDYTGK-FTR-SDKEGPDYEHVINA 674  
QY 580 MINSNQIQVGADGITFTFIDISNKPAGIENTRITRBDIGFANNTGSLDANKPRLPTG 639  
Db 675 AKAK-KVVLKFKKMDASLADYSNVLVKINDTLQTLSEDAVTLVSNDATVITTEFT 733  
QY 640 INAGKEILNVSAINPAINSGQLDFPMRLSTANTKESGSAITLIDLYNSQVPLTRAGD 699  
Db 734 IKGDVVEASGKALSG-SGKVVNELYMGVYKDTSGVHKKFNSENK-IT-LSSTST-P 789  
QY 700 T-GPNVTKLGEILKVGKGTADLTCKNIGVADSDNSLTJYKLAKLTSLDLDLVMTKT 758  
Db 790 LKLA-KIDQVAKYATLVDKRYKVFSTVINSAA-ANA-FISESKT-DS--IOVNG 843  
QY 759 LTAASDVAIVD-SGNTATKQNGDLTFESQNTGAPPAINSKIYVDGKFTDNGGIALDG 817  
Db 844 TSTVYVKEDEINTNASDLDL-KVALSK-LVDIAGNE-STNNTQPLATKAGINLDSVAPV 900  
QY 818 TTYIT-KDKVGFAGDGLSKSKPYLDKDKLKGVEVELTTGGINAGKAITGLSNTLTD 876  
Db 901 VVGEPEV-VDEKITTFTESENLTVSISGEVLTSTFTVTVSDNKLAIKDY-SVAIANNQ 958  
QY 877 TNATGHTVQLGIV-DSYDKTRASIGDVLNAGFNLNKNGDAKDF-VSTYDVFINGNA 934  
Db 959 VYITLS-DNREYATAYKATKANAKL-ITDONGDKKNALADPTKTKATKVEASGLISDAA 1016  
QY 935 TTAKTYTGKSKVAIVDVNVTGITHLTGADGNK-NOIG-VKTTTLTKTKAGDKAINFS 992  
Db 1017 KTNLNEITKAKDAKATGTEGTAATNOIVGSKDALQVAIDVAVELVKNDFTATLQOLTDAK 1076  
QY 993 -VNSGDDKALI-NAKDIADNNTLAGIEIRNTKGTADALQTFQ-VKVKVNGGDD-NDAD 1048  
Db 1077 TDLTAITAYNAKVEDLSILVAPDVLGTTDNGITTGAVAGTGETLKYVTSOSAN 1133  
QY 1049 T-TIVGKDA-KTNOVNTLK-LKGRNGDIDQTNKGGTGT-FGINQSGKLKAGNNNTIL 1101

RESULT 6  
ENTRY B43855 #type complete  
TITLE high-molecular-weight surface-exposed protein - Haemophilus influenzae  
ORGANISM #formal\_name Haemophilus influenzae  
DATE 10-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Nov-1994  
ACCESSIONS B43855  
REFERENCE B43855  
#authors Barenkamp, S.J.; Leininger, E.  
#journal Infect. Immun. (1992) 60:1302-1313  
#title Cloning, expression, and DNA sequence analysis of genes encoding non-typeable Haemophilus influenzae high-molecular-weight surface-exposed proteins related to filamentous haemagglutinin of Bordetella pertussis.  
#cross-references MIMD:92192797  
#accession B43855  
#status Preliminary  
#molecule\_type DNA  
#residues 1-1477 #label BAR





```

#journal      Mol. Microbiol. (1989) 3:517-539
#title        Analysis of the yopA gene encoding the YopI virulence
              determinants of Yersinia spp.
#cross-references EMBL:89343638
#accession    S04912
#status       preliminary
#molecule_type DNA
#residues     1-455 #label SKU
#cross-references EMBL:X13882; NID:g48606; PID:g48607
GENETICS
#genome       plasmid pY6471/76
SUMMARY
#length 455 #molecular_weight 47136 #checksum 6266

Query Match      1.6%; Score 209; DB 2; Length 455;
Best Local Similarity 34.9%; Pred. No.2,156-07;
Matches 44; Conservative 33; Mismatches 45; Indels 4; Gaps 4

Db 55 PPYPAGGLMSAKIHSTAGATLEAAKGAAYAGASITGVNSVAIGPLRALGDSA 114
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 1825 PVOGQRNIDSSASKSHVAIGFOAKAGAEVAVALGRQT-QAGNOSTIGNMAATGPOS 1883
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 115 VYGGAST-AQDGVAGIARAST-SDTGVAVGFNSK-ADAKNSVAIGHSSVAANHGYSI 171
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 1884 IALGTGNVYAGHSGAIDPSTVRADNSYSGVNNQFTDAQTIDYFGVGNNTITESNSV 1943
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 172 AIGDRS 177
      |::|
Oy 1944 ALGSNS 1949

RESULT 9
ENTRY S70791 #type fragment
TITLE vsab protein - Mycoplasma pulmonis (SGC3) (fragment)
ORGANISM #formal_name Mycoplasma pulmonis
DATE 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Sep-1997
ACCESSIONS S70791
REFERENCE S70790
#authors Bhugra, B.; Voelker, L.L.; Zou, N.; Yu, H.; Dybvig, K.
#journal Mol. Microbiol. (1995) 18:703-714
#title Mechanism of antigenic variation in Mycoplasma pulmonis:
interwoven, site-specific DNA inversions.
#accession S70791
#status preliminary; nucleic acid sequence not shown
#molecule_type DNA
#residues 1-555 #label BHU
#cross-references EMBL:023947; NID:g780764; PID:g780766
GENETICS
#gene vsab
#genetic_code SGC3
#map MARX #length 555 #checksum 3948

Query Match      1.6%; Score 217; DB 2; Length 555;
Best Local Similarity 19.9%; Pred. No. 3,496-08;
Matches 104; Conservative 146; Mismatches 247; Indels 26; Gaps 26;

Db 21 KPMIDKSGNSGDQKGNKNNAGDANAGEMKNNAGDANAGEMKNNAGDAN-GEKKNNA 79
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 1531 KSVINKE-QVANDNA-KQGINEDNAFYKGEKASDNKTKAAVTVDLNAVQTPLTFA 1578
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 80 GDANAGEMKNNAGDANAGEMKNNAGDANAGDGMKNNAGDAN-AGEKKNNAGDANAGD 138
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 1579 GDTGTAKKAGETLTIKGGQIDTKLTDNNITGVAAGTDGFLYKLAKPLTINLSVNAAGTK 1638
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 139 MGNKNNAGDANAGEMKNNAGDAN-G-EMGKNNAGDANAGEMKNNAGDANAGEMKNNNA 196
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 1639 IDDGVSFTVSDSSGQAKANTPVLASANGDLDGKGVTSNKGKGTFTDANVVOOLNEDV-ENLL 1697
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 197 GDANAGDMKNNAGDANAGEMKNNAGDANAGEMKNNAGDANAGANAGEMKNNAGD-ANAG- 254
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy 1698 GLGNAGN-DNADGNOVNTADT-KKDPNPGSSSSNTYVIKAGTVLGG-KGNNDTEKLATGGV 1754
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 255 EMKNNAGDANAGEMKNNAGDANAGEMKNNAGDANAGEMKNNAGDANAGEMKNNAG 314

```

Qy	1755	QGVGVKDKDNAN-GDLSINWVWTKQDGSK-KALLLTATYNAAG-TVYLTINNPFAEIDRINEQ	1811
Db	315	DANAEEMKNNAGDANAEEMKNNAGDA-NAGEEM-KNNADANAEEMKNNADANAGE	372
Qy	1812	GIRFFHVNDGNOEPYVQGRNGIDSSASKSHVAIGFOAKKADGEAAVAIGRTQAGNOSIA	1871
Db	373	MGRNNAGDANAG-EMKNNAGDA-NAGEEMKNNAGDANAG-EMKNNNA-GDANAGEM-GK	427
Qy	1872	ICDNNQATGDGDSIALGTGNVVAKKHSGAIGPSTVYKADNSVSGNNNOFTDATQTDVFGV	1931
Db	428	-NN-A-GDANAEEMKNNAGDANAEEMKNNAGDANAEEMKNNAGDANAEEMKNNAGD	484
Qy	1932	GNNITVTSNNSVAAGLSNAISAGHAGTAAKSSGRTAGTTTACATGTYKGFAGQTAVGA	1991
Db	485	ANAGEEMKNN-AGDANAEEMKNNAGDANAEEMKNNAGDANA	526
Qy	1992	VSVGASGAERRIQNVAAGEVSAATSDAVNGSQLYKATQIGIANA	2034
RESULT	10		
ENTRY	140711	#type complete	
TITLE	sapB protein - Campylobacter fetus		
ORGANISM	#formal_name Campylobacter fetus		
DATE	16-Aug-1996	#sequence_revision 16-Aug-1996	#text_change 16-Aug-1996
ACCESSIONS	140711		
REFERENCE	A56999		
#authors	Dworkin, J.; Tummuru, M.K.; Blaser, M.J.		
#journal	J. Biol. Chem. (1995) 270:15093-15101		
#title	Segmental conservation of sapA sequences in type B Campylobacter fetus cells.		
#cross-references	MolDb:95318069		
#accession	140711		
#status	preliminary; translated from GB/EMBL/DBJ		
#molecule_type	DNA		
#residues	1-936	#label RES	
#cross-references	EMBL:025133; NID:g801999; PID:g802000		
GENETICS			
#gene	sapB		
SUMMARY	length 936	#molecular_weight 96310	#checksum 6995
Query Match	1.6%;	Score 210;	DB 2; Length 936;
Best Local Similarity	20.3%;	Pred. No. 1,72e-07;	
Matches	134;	Conservative 183;	Mismatches 300; Indels 44; Gaps 41;
Db	213	ENGTTYVNGTNGKSTLSAFDRKIDGAGARDT-LNA-IFTAN-NNAAATKL-D-OTEIFDKS	267
Qy	1150	FAGTNG-GSLDKSRPHLSK-DGINAGKKTITNGEIAIIONSNDPAVGTGKIYDLKTELENK	1207
Db	268	VKGVTYN-EN-INIISDLTSGD-F-VPNGEKKGENVLDIASFAIDAKSVNVEITGT	323
Qy	1208	ISTTAKTAQNSLHEFSVADEGNNFTVSNPSSYDTSKTSBDVTTFEAENSIITTKVNGGV	1267
Db	324	ITTAFTAA-G-TG-YVDVAGKRTALTRADATSVLRAFTNDFTITSANATSVNLKTSGA	380
Qy	1268	RVGIDQTFGLTTPRLTYGNNNGKRIVIDSONQN-TITGLSNTLIANTYNDKG-SVRTTEQ	1325
Db	381	AKSAT-ITSANAAKNITITDATGVAAVTSATRAVENLVKHAFTNVTLAGNMDKLATVTLDNA	439
Qy	1326	GKIIKDEKDTKRAASIVLVLSAGFMLOENGEAVD-FVSTYDT-VNFADGNATTAKTAYDDT	1383
Db	440	ALTAADIKSASTLINLINSVYNGNISTAKKADVIVHLSGSAAYKUNTTATATQGYTLKA	439
Qy	1384	SKTSKY-V-YDVVDDDTTIEVKDKLLVKTTLTTLSTGTG-ANKFALSNQATGALVKAASD	1440
Db	500	NATP-NSLEPDSFGTAKTTSVYASGSGRTLVIKGEVETLVNIDTTFANGADVSPFGDAQ	558
Qy	1441	IVAHNLNTLSGDIQAKAGASQANSAG-YVDADGKV-IYSTDNKKYQAANDGTVDKTK	1498
Db	559	SGKSVKGTGIDDKIEFVYGTLLTEGSIYIDGANDTITAMKSAALTSANFTMIKTIENVAIS	618
Qy	1499	VAK-D-KLVAAQAGPPDGLQAMNNKTVYINKEOVYNDANK-OGINEN-APVKGLEKRAASD	1554

Db 619 DAVATAADLSSAFKNIITITRE-AADTTLTINK-DOYINFAADROSVKLTITKLDVYIG 676  
1555 NKTNAAVTGDLDNAVQOTPLTFAKLTGKLTITKGGOTDITKIDNNIGVAG 1614  
Db 677 AND-VKIVIVDAAKKASIALGFEATDKALV-IDTGIELTITSLYKATSPENTANTV-N 733  
1615 TDGFTVRLAKDLTNLNSVNGKIDKIDKVSFVDSGQAAANTPVLSANGLDIGKVIYN 1674  
Db 734 -AKLT-DVTSIITD--GDAK--ITLGHAGTAGTDISKVSMIDAR-ALKAGLTIDASAATL 786  
1675 VGRKTKTDANVOQLNEVRRLGLGNAGNMDNQVNIADIKDPNCGSSSRVYKA 1734  
Db 787 GANATIKGSGADSIYKGNIVVDL--VAGGD-DITTLKGRKKTITTVNNFNGDKI 843  
1735 GTVLGKGNNDTEKLAGGVGVYDKGNMNGDLSNVWVKOKGSKKALLATYNAQOT 1794  
844 D 844  
QY 1795 N 1795

RESULT 11  
ENTRY high-molecular-weight surface-exposed protein HmW1 -  
TITLE Haemophilus influenzae  
ORGANISM #formal\_name Haemophilus influenzae  
DATE 10-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Mar-1998

ACCESSIONS  
REFERENCE A43855  
#authors Barenkamp, S.J.; Leininger, E.  
#journal Infect. Immun. (1992) 60:1302-1313  
#title Cloning, expression, and DNA sequence analysis of genes encoding nontypeable Haemophilus influenzae high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of Bordetella pertussis.

#cross-references M01D:92192797  
#accession A43855  
#status preliminary  
#molecule\_type DNA  
#residues 1-1536 #label BAR  
#cross-references GB:U08876; GB:M64616; NID:9475770; PID:9475771  
#note sequence inconsistent with the nucleotide translation sequence extracted from NCBI database (NCBI:89235, NCBI:89239)

SUMMARY #length 1536 #molecular-weight 159916 #checksum 3096

Query Match 1.6%; Score 217; DB 2; Length 1536;  
Best Local Similarity 18.7%; Pred. No. 3,49e-08;  
Matches 154; Conservative 242; Mismatches 372; Indels 54; Gaps 51;

Db 668 LNVSEGEFNLITDRSGSDACTIOPVNLNGISFNKDTTFVFNARVNPDIKA-PIGI 726  
QY 1282 LVYGNNGNGIYIDSONGNTIT-GLSNLAVNTDKGVSRTEDGKIIXDDKIRASI 1340  
Db 727 NKYSILNVA-SFNGN-IS-VSGGGSVDFT-LIASSNNVOTPEGVINSKY-ENVSTGSSIR 781  
QY 1341 VDLISGFNLQNGEAVFSTYDVNFADGNATKATYDDTSKTSKVYVYVNDT-T 1399  
Db 782 FRTSGSTKGFSEKDLTINANGNTITLQVET-DGM-LGKIYAKKNIPEGG-NITF 838  
QY 1400 IEVKD-KRLGVKT-TTLTGTGANKFALSNGATGALYKASDIYAHNL-TLSGDIQIAK 1456  
Db 839 GSRKAVTELEGVNTINNANVTLLIGSDPDNHQ-KP-LTIKQVLIINSGLTLAGNIVNIA 896  
QY 1457 GAGQANSS-AGYVDADGN-KVIYDSTDKYQAKNDGTVDKTEVAKKDLVAQAQPDGT 1514  
Db 897 -GNLTVESNANFKATINFTFNNGGLFDNKG-NSNISIAAGAFKIDISK-NLSITINS 953  
QY 1515 LAQMNKSVIN-KEQVNDANKKOGINEDNAFVGLKKAASDNKNAAVTDLDNAVQOT 1573  
Db 954 SSTYKTIISGNTINKNGD-LNITNBSDTMEMQIGDVSQKKEGWLITSSDKINITKQITIK 1012

QY 1574 PLTF-ASDGTAKKLGELITIKGGQDTNKLTIDNNIGVAG--TDGF-TVYLAKDITNL 1629  
Db 1013 AGVDGSDSDATNANNTITTEKL-TODLNISGEN-KAEITAKGSDLTIGNTISAD 1070  
QY 1630 NSVNAAGTEKID-KVSVFVDSGQAKANTPVLSANGDLGKVSNNGKTRDTDAANVQ 1688  
Db 1071 GNNAKVFENQYKDKISADGKVTLSKVEFSGNNNTEDSDNNAGLITDAKNVTNN 1130  
QY 1689 QINEVKNLGLGNAGDNBDNQVNI-ADIKDPNCGSSSRVYKATVJLGGKNNDE 1747  
Db 1131 NITSHKAVISATSEIITKGTITNATGNVEITAGTSLIGTSSGSVTLTATEGA 1190  
QY 1748 KLAT-GGVGVYDKDGNANGDLSNVWVKOK-D-GSKK-ALLATYNAAGTNTLNNPAE 1803  
Db 1191 LAVNSIGNTVTV-TANSALTLTLAGSTIKTEESTYSSGSDTG-CTISG-GVEY-K 1245  
QY 1804 -AIDRINEGIRFFHYNDQENPEVYQGR-NGIDSSASGSHVAIGFOKADGEAVALGR 1861  
Db 1246 ATESLTQSNK-IRA-TTGEANVTSATGTI-GGTISGNTV-VTANAGD-LTVNGAEI 1300  
QY 1862 QTOA-GNOSTAIGDNQATGDSIAGTGNVYAKGSHGALGDPSTVKADNSVSGNNQF 1920  
Db 1301 NATEGATLITSSGKLTTEASSHITSAGQVNLSDQDSVAGSINAANVTLTGTLTV 1360  
QY 1921 TDATQDVFVGNNITVETESNVALGSNSAISAGTHAGTQAKKSGTAGTTTATAGTGV 1980  
Db 1361 KGSNINATGTLVINKADE--LNGAALGNHVVAANANGSGSVIATNSRVNTGDLI 1418  
QY 1981 KGFAGQTAAGAVSAGSAGERRQVNAAGEVSATSDAVNQS-QLYKATQIGIANTNEID 2039  
Db 1419 -TINGLITISKNGINTVLKGVKIDVKTQIPGIASVDEYIEA 1459  
QY 2040 HRHONENKANAGCISAMAMASMPQ-IT-PERSMTGTGSIAT 2079

RESULT 12  
ENTRY S04911 #type complete  
TITLE yopA protein - Yersinia enterocolitica virulence plasmid  
ORGANISM #formal\_name Yersinia enterocolitica  
DATE 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Sep-1997

ACCESSIONS  
REFERENCE S04911  
#authors Skurnik, M.; Wolf-Watz, H.  
#journal Mol. Microbiol. (1989) 3:517-529  
#title Analysis of the yopA gene encoding the yopA virulence determinants of Yersinia spp.

#cross-references M01D:89343638  
#accession S04911  
#status preliminary  
#molecule\_type DNA  
#residues 1-422 #label SKU  
#cross-references EMBL:X13881; NID:948609; PID:948610  
#note the authors translated the codon CCA for residue 282 as Thr

SUMMARY #length 422 #molecular-weight 44134 #checksum 8814

Query Match 1.5%; Score 206; DB 2; Length 422;  
Best Local Similarity 22.6%; Pred. No. 4.23e-07;  
Matches 60; Conservative 81; Mismatches 115; Indels 10; Gaps 9;

Db 126 AIGARATSDGVAVGNSVYDAKNSVAIGHSSHVAVDHDYSAIGDRSKTDRKNSVIG 185  
QY 1832 GIDSSA-SGRHSAVIGFOAARADGAAVAIGQTOAG-NO-STAIIDNQAIGDQSIAG 1887  
Db 186 HESLNQTLHAGTKDQDAVNAVOLKEIEKTQVANKKSAEVLGIANNYTDKSAETL 245  
QY 1888 TGNVYAG-KH-SGALGDPSTVYKADN-SYVGNNQNFDAQTOGVFGVGNITVETES-NSV 1943  
Db 246 ENARKEAFDLSNDALMAKSHNSVARTTLETAEHPNKSAAETLARANYADSKSHL 305

Qy	1944	ALSSNSAISGATGACGAAKSSDCACTGTTAGATGTVGVGFAQCTAVGAVSVGASGAEBERI	2003
Db	306	QT-ANSTYDVTYVNSSTKKA-IRSNQYTHKRRQLDNRDLKDTYVDKGLASSAALNSLF	363
Qy	2004	QNVAAEEVATSTSDYANGSQLYKKAQIGANATNATNELDRIHONENKANAGISSAMAMASMP	2063
Db	364	QPYGVKVNFTAGVGGYRSSQALAIg 389	
Qy	2064	QAVIPGRSMVTGGIATHNQGAVAVG 2089	
RESULT	13		
ENTRY	S04534	#type complete	
TITLE	invasin precursor - Yersinia pseudotuberculosis plasmid pIBB		
ORGANISM	#formal_name Yersinia pseudotuberculosis		
DATE	20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 06-Sep-1996		
ACCESSIONS	S04534; S04910		
REFERENCE	S04534		
#authors	Rosqvist, R.; Skurnik, M.; Wolf-Watz, H.		
#journal	Nature (1988) 334:522-525		
#title	Increased virulence of Yersinia pseudotuberculosis by two independent mutations.		
#cross-references	MUID:89302441		
#accession	S04534		
#status	preliminary		
#molecule_type	DNA		
#residues	1-434 #label ROS		
#cross-references	EMBL:X12758; EMBL:X13883; NID:g48639; PID:g48640		
REFERENCE	S04910		
#authors	Skurnik, M.; Wolf-Watz, H.		
#journal	Mol. Microbiol. (1988) 3:517-529		
#title	Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia spp.		
#cross-references	MUID:89343638		
#accession	S04910		
#status	preliminary		
#molecule_type	DNA		
#residues	1-434 #label SKU		
#cross-references	EMBL:X12758; EMBL:X13883; NID:g48639; PID:g48640		
GENETICS			
#gene	yopA		
#genome	plasmid		
SUMMARY	#length 434 #molecular_weight 45054 #checksum 7718		
Query Match	1.5%; Score 207; DB 2; Length 434;		
Best Local Similarity	24.7%; Pred. No. 3,38e-07;		
Matches	65; Conservative 80; Mismatches 107; Indels 11; Gaps 10;		
Db	142	SKALDSDAVTQYGASSTADKDG-VAIGARASDVGAVGVGNSKVDQNSVAIGHSHVAA 200	
	1835	SSASGSKHVAIGFOAKKAGCEAAVAIGRTQAGNOSIAIGDQAQATGGOSTAIGTVAVAG 1894	
Db	201	DHGYSIAIGDHSKTDRENSVSIAGESLNRLTHLAAGTEDTDVAVNAQLKENAETLENA 260	
Qy	1895	KHS-G-AIGDPSYKADNSVGVGNH-N-OPTD-ATQDFGVGNNTTV-E-SNSVALG 1946	
Db	261	KRETLAOSNDVLDAKKHNSVARTTLETAEEHNNKKSAAALYSA-KVYADNSSTITLKT 319	
Qy	1947	SNSATSACTHAGTQAKKSDGTAGTTTGAAGATGYKGAGGTAAVGAASVSGASGERIQQY 2006	
Db	320	ANSTYDVTYVSSSTKKA-ISESNQYTHKRRFOLDNRDLKDTYVDKGLASSAALNSLFQY 378	
Qy	2007	AAGEVSATSTDAVNGSQLYKATQCIANATNELDRIHONENKANAGISSAMAMASMPQAY 2066	
Db	379	GVGKVNFTAGVGGYRSSQALAIg 401	
Qy	2067	IPGRSMVTGGIATHNQGAVAVG 2089	
RESULT	14		
ENTRY	UMMS	#type fragment	
TITLE	period clock protein mouse (fragment)		

ORGANISM	#formal_name Mus musculus #common_name house mouse			
DATE	31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change			
ACCESSIONS	A24403			
REFERENCE	A24403			
#authors	Shin, H.S.; Bargiello, T.A.; Clark, B.T.; Jackson, F.R.;			
#journal	Nature (1985) 317:445-448			
#title	An unusual coding sequence from a Drosophila clock gene is conserved in vertebrates.			
#cross-references	MUID:86014384			
#accession	A24403			
#molecule	type DNA			
#residues	1-713 ##label SHI			
#cross-references	GB:X02966; GB:M12039; NID:955125; PID:e7759;			
COMMENT	PID:g1334150			
Mutations within the per locus of the fruit fly affect a variety of natural biological rhythms (from long-term circadian to short-period behaviors). The genome of the mouse contains a homologous locus with multiple tandem repeats of nucleic acid hexamers (ACNCGN, TCAGGC) that encode poly(T/S-G) tracts up to 48 residues long.				
COMMENT	The serine residues of the S-G repeats found in certain proteoglycans are attachment sites for glycosaminoglycans, although it is not known whether this protein binds proteoglycan.			
CLASSIFICATION	#superfamily period clock protein			
KEYWORDS	circadian rhythm; tandem repeat			
SUMMARY	#length 713 #checksum 3504			
Query Match	1.5%;	Score 207;	DB 1;	Length 713;
Best Local	Similarity 18.4%;	Pred. No. 3.38e-07;		
Matches	96;	Conservative 131;	Mismatches 277;	Indels 18; Gaps 17;
Db	98	TTTGATGATGCTGTGGRKATGRGRTGDDPDGVTYA--RAIVTARVATGRTGATVTE	155	
Qy	1510	TPDGLIQQMNKYSVINEQYVNDANKOGINEDNAPYKGLERKASDNKKNAAYVGDINA	1569	
Db	136	TGTAKVYDTC-TGTGTAKTGTAKVTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTDR	214	
Qy	1570	VAQPPPLTFPAGGCTTAKKLEETVLINGGOTDYNKLTLDNNIGVAATDFEYVFLAKDLTNL	1629	
Db	215	GTGTGTGTGTGTGTGTGTGTGTAKVATGTAKVATGTGTGTGTGTGTGTGTGTGTGTGT	274	
Qy	1630	NS-VNAG-GRKID-DKGVSPVDSGSAK-ANPVLVSANGDLGKVISNNKGT-KDTDA	1684	
Db	275	GTTAKVYTGTTGT	333	
Qy	1685	-ANYQOLEVFNLLIGLGN-AGNDNADGNQVNIADIR-KDPNSGSSNNRTVIAKAGTVLGR	1741	
Db	334	TGTGTAKVTTGT	393	
Qy	1742	GNNDTEKLAGVQVGVYKDD-GNANDJLSNVMWYKQDKSGKALLATYVMAQOYNYLYNN	1800	
Db	394	SGSGTAKVTTGT	453	
Qy	1801	PAEADIRINQGIREFPHVNDGQEPVQGRNGIDSSASGKHSVAIGFOAKADGEAAVAIG	1860	
Db	454	TGPDGTAKVTTGT	512	
Qy	1861	KQTDAGQOSTAIDNMQATGDSIALIGTGNVYVAGKHSALIDPSTVADADNSYSGNNQD	1920	
Db	513	AKVYGTDT-ETAK-VTGT	569	
Qy	1921	TDAGQDIDFEGVGNLITVESNSVALGNSAISAGIHAQTQAKKSDGTAGTTTGAATGIV	1980	
Db	570	TGLGSGSGSTAKVTTGTAK-VTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	610	
Qy	1981	KGFAGQTAGAVSVGASGAERRIQWVAAGVEASVSTDAVNGS	2022	
RESULT	15	A48995	#type complete	
ENTRY	paracrystalline surface layer protein RsaA - Caulobacter			
TITLE				







```
QY 724 LTKNIGVAASTONSLTVKLAKTSLDDAVNTKTLASDKVTYVSGNNTALQNGDLTF 783
D 228 ATLGAVKATKATTKLTNAASVLTLTNAAVLGAIDNTGGDNVGLNGLALSOVTDI 287
QY 784 SKONNGATPATSKTIYVDGKFLDNNIGIALDGTITIK-DKVEFAGQDGLSKRPL 842
D 288 GNTN-SLATISVAGATATLGAVIKATTKLTDAASAVK-FTNPVVTGAIDNTGNANN 345
QY 843 DKDKLKEVEETITNGINAGGKAITGLSLTLDATNATGHTQIGIYDSTDKTPAASIG 902
D 346 IYTFEG-NSTYTGNGN-TNALATVN-VGAGLQYOGV-VKANTINLTNASTATFNP 401
QY 903 DVLNNGFNLKNNGDADKDVSTYDVTDFINGNATTKATYVDGASVAVADVNDGTIT-HL 961
D 402 VVTGAIDNTGNANGIY-T-FTGNSVYV-G-DIGNTALATV-NVAGATATLGAV-I 454
QY 962 TGADGNKNOIGYKTTTLTKTDKAKKALNFVNSGDDKALIAKDIADNLTLAGBIKRT 1021
D 455 KATT-TKL-T-NAASVL-TLTNAAVLGAIDNTGGDNVGLNGL-ALS-QVT--GN 505
QY 1022 KGTADTALOTFQVKKYKENGDDNDADITTVGKDAKTNOVNTLKLKKGKGLDIQTNKDG 1081
D 506 IG-NTNSLATISVAGAT-TLGAVIKATTK-LTD-AASAKFT--NPVVTGAIDNTG 559
QY 1082 VTEGINTSGKAGNNTLTNNGLSIKNTAGNEQIOVGADGVKAKVNGVAGAGIDGT 1141
D 560 NANGIYVFTGNSVYTGIDGNSTLATSISVAGATATLGAVIKATTKLTNAASVLTLTN 619
QY 1142 RTREIFGAGNG-SLDSKRP-LSKDCINAGGKKTITIOGELAONSDAVTGKTI 1199
D 630 ANAVLTGAIDNTGGDNVGLNGLALSOVTDIGNTSLATISVAGATATLGAVIKAT 679
QY 1200 LKTELENNISSAKTAQNSLHFEVSADQ-GNFTVSNPYSYDTSKTDVITFAGENG 1258
D 680 TTKINNAVASAKFTNPVVTGAIDSTGNAN-NGIYV-T-TGNSYTGIDGNNTALATV 736
QY 1259 TTKVKKGVVRGIDQTKGLTTPKLTGVNANGGIYIDSONGNTITG-LSNLTAVNTDK 1317
D 737 GAGTATLGAVIKATTKLTNAASVLTLTNAAVLGAIDNTGGDNVGLNGLALSOV 796
QY 1318 GSVRTBEQKIHKED-K-TRASTIVDLASQFNLOGNEADVSTYDVAFAG-NAT 1374
D 797 TGDIGNTSLATISVAGATATLGAVIKATTKLTNAASVLTLTNAAVLGAIDNTGG 856
QY 1375 TAKVYDDTSKTSKYVYDVN-VDDTIEVKDKKLGKTTTLTSTGTGANKFALSQATGD 1433
D 857 DNVGLNGLALSOVTDIGNTSLATISVAGATATLGAV-I-KATTKLTNAASVLT 914
QY 1434 ALVKASDIVAHLNTLSGDIQTKAGSAQANSSAGYVADGNKVLYDSTDKRYOAKNDGTV 1493
D 915 TNANAVLTGAIDNTGGDNVGLNGLALSOVTDIGNTSLATISVAGATATLGAVIK 974
QY 1494 DKTRVARDKLVAAQOTPD-GTLAOMNKSVINKBOVDAN-KKQGINDNAPVGLK 1551
D 975 ATTKLTDAASAVKFTNPVVTGAIDNTGNANGIYVFTGNSVYTGNGNTALATV 1034
QY 1552 ASDNKTNAAVYVDLNAVAQF-PLTFAGDTGTAKKILGETLTIKGGQDINKLIDNNIG 1610
D 1035 AGLLOVGVYK-ANTI-NLTNMAAVFTNPVVTGAIDNTGNANGIYVFTGNS-IT 1091
QY 1611 V-VAGTDTGVYKLDNLNLNLSVNGAGTIDDKVS-FVDSGGAKKANTPVISANGLDIG 1668
D 1092 GNVGNTALATVNGAGLLOVGVYKANTINLTNMAAVFTNPVVTGAIDNTGNANN 1151
QY 1669 GKVINNVGAGTIDTAANQ-OLNEVR-NLGLGNAGNDAGNOVNIADIKKDRSGSS 1726
D 1152 GIVFTGNSVYTGIDNTNA--LATVNGAGI 1181
QY 1727 SNRTVIKAGTIVLGKGNNDTEKLATAGVQGV 1758
```

RESULT 2

```
ID SLAP_BACSH STANDARD: PRT: 1176 AA.
AC P38537;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DE 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DT SURFACE-LAYER 125 KD PROTEIN PRECURSOR.
OS BACILLUS SPHERICUS.
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-2362;
RX MEDLINE: 89327128.
RA BOWDITCH R.D., BAUMANN P., YOSTEN A.A.:
RL J. BACTERIOL. 171:4178-4188(1989).
CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.
CC -1- SUBCELLULAR LOCATION: CELL WALL.
CC -1- SIMILARITY: CONTAINS 3 COPIES OF THE S-LAYER HOMOLOGY (SLH)
CC DOMAIN.
DR EMBL: M28361; G556012; -.
DR PIR: A33856; A33856.
DR HSSP: P31417; IMDC.
DR PROSITE: PS01072; SLH DOMAIN: 2.
KW SIGNAL; CELL WALL; S-LAYER; REPEAT.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 1 1176 SURFACE-LAYER 125 KD PROTEIN.
FT DOMAIN 32 91 SLH 1.
FT DOMAIN 92 151 SLH 2.
FT DOMAIN 152 210 SLH 3.
SQ SEQUENCE 1176 AA; 125225 MW; 0C26885C CRC32;

Query Match 1.8%; Score 243; DB 1; Length 1176;
Best Local Similarity 20.7%; Pred. NO. 3.22e-13;
Matches 186; Conservative 228; Mismatches 426; Indels 57; Gaps 55;

D 268 VYTTATKVSATNKLKVVVEFDCTVDKETAEDANALALSGTIKTSVLAANKTKATV 327
QY 231 LTATKAASSIATVGS-NAQALGFATVAGSTOVNLNGLALGSGOVLQKDNVDNANV 289
D 328 TLTDKLNKKNKADALSIISNVKAGDEI-NVKNVEFT-AVDNRIPEYTEVKSIGTRAVYT- 384
QY 290 RAYAPDDNOLPNRKAFAFKNGATDVESIGNSNGDSIRKRIINVGASA-DTAVVNAQ 348
D 385 LSEPVENISSTNFTIDGATFNGVYMGAKNTVILTPYSSALSVDGKLTYSKAPAG 444
QY 349 LKEAV-RLANQIIEFGGDSNNRVEKGLG-KTLITIG-GAOTFALTDHNISGVONGSLK 405
D 445 FVSLNST-HEKVV--ED-KKAPVTEATATLENTLTFSSDIDMDVYKASNYWKSQDS 500
QY 406 -VQLETLTSLKMTTEENTLANKETVGTKTLTDTKIGFT-N-DWNGIDSKPYLDK-DT 461
D 501 KKEASEFERIADNKYEFKSEKTLPTGKYDVYED-IKDYSNKAIAKDKRYVTPRID 559
QY 462 GHNAGG-QKIRKLAVGVDDDAATYGOLAKVYNGQTAESLQGFYKVKVNDKNDANDSKII 520
D 560 QTRPEVRKVTALDEKTIKVTFSKYVDSAIK-TGNVTAKDKDKVSVYKYVDSKDS 618
QY 521 -TVGRKNNRPDQVNTLKLKENGVDVTTENGVTFEGLNONGIIVGNSSTLNDGSLVK 579
D 619 SVIIDLYKSVAGENTIN-INKVDAATKLNTM-LDYIK-FTR-SDKEGDEHYINAD 674
QY 560 NNSNKQIOVGADGITTFIDISNKPAGAGIENTTRITRDIGIFANNNGSLDNKRPILPTG 639
D 675 AKAK-KVILKFPDKKADASLADYSNVLKINDTLQTLSEDAVLISVSDATVITTEAET 733
QY 640 INAGKEILTINQASINPRTNGGOLDPMNRRLSTANKEKGSATIKDLNLNOVPLTFEGD 699
D 734 IKGDVFPASGKALSG-SKAVNVNELQVYKDYSGVNHKFGNSENK-IT-LSSTST-P 789
QY 700 T-GPNVTKKLEILKVKGGKTADDLTKNNIGVADSDNSLTVLAKTLSDLDLAVNKT 758
D 790 LKLA-KIKKQVDKATLALVDKRYKVFSTVINSAA-ANA-FISESHKI-DS-IQVNG 843
```



OY 759 LTASDKVTVD--SGNNTAKLONGLDLEFSKONTGAPATNSKITYGVGLKFTDNNNGIALDG 817  
 DB 844 TSTVAVKREDEINTNASDDL-KVNLK-LVDIAGNE-STNNTPIAKAGINLIDSVAPV 900  
 OY 818 TTYIT-KDVGAKADGSLDKSKPTLKDCKLVEVEYETITNGINAGGALIGLSSTLIDA 876  
 DB 901 VGEDEV-VDEKETITFESENLSVSIGEVLSDTFTVTRVSDNKDLAIDY-SVALANNQ 958  
 OY 877 TMAITGHVTLQGLV-DSTOKTRASIGDVLNAGFNILKNNGDAKDE-VSTYDVFDEINGNA 934  
 DB 959 VVITLS-DNREVAATAYKATAKNAKL-ITDDNDCKNALADEFKTATAYEASGTLSDAA 1016  
 OY 935 TTAQVYDCKASRVADVVDGTTIHLTGADGNK-NOIG-VKTTTLTKTDAGDKRAINFES 992  
 DB 1017 KTNLNNEITKADAKATGEGTAATNOIVGSKDALOVAIDVAVELVKNPDAATLQOGLTDAK 1076  
 OY 993 -VNSGDDKALIT-NADIANLNTLAGEINTGTADTALQITQ-VKAYENDDDD-NMAD 1048  
 DB 1077 TDLTAITAYNAKVEDISLLVAPDLVLTGTDNGITINGVAGTGLTKVTSDAAN 1133  
 OY 1049 T-ITVGKDA-KTNQVNTLK-LKGNGLDIOGNKDGTVI-FGINTOSGLKAGNNTLIN 1101  
 RESULT 3  
 ID HLVA\_PROMI STANDARD; PRT; 1577 AA.  
 AC P16466;  
 DT 01-AUG-1990 (REL. 15, CREATED)  
 DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)  
 DE HEMOLYSIN PRECURSOR.  
 GN HEMA.  
 OS PROTEUS MIRABILIS.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 CC ENTEROBACTERIACEAE.  
 [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.  
 RC STRAIN-ISOLATE 477-12;  
 RX MEDLINE; 90170827.  
 RA UPHOFF T.S., WELCH R.A.;  
 RU J. BACTERIOL. 172:1206-1216(1990).  
 CC -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD  
 CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY  
 DEFINED.  
 CC -1- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM  
 ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HEMA  
 REQUIRES HEMB FUNCTION.  
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
 CC -1- THE CONSERVED AMPEPTATHIC DOMAINS IN SHLA AND HPMA MAY BE  
 RESPONSIBLE FOR PORE FORMATION.  
 CC -1- SIMILARITY: TO S. MARCESENS HEMOLYSIN (SHLA).  
 DR EMBL; M30186; G150890; -.  
 DR PIR; A35140; A35140.  
 KM HEMOLYSIS; TOXIN; OUTER MEMBRANE; SIGNAL.  
 FT SIGNAL 1 29  
 FT CHAIN 30 1577 HEMOLYSIN  
 SQ SEQUENCE 1577 AA; 165869 MW; 979B0608 CRC32;  
 Query Match 1.7%; Score 221; DB 1; Length 1577;  
 Best Local Similarity 17.7%; Pred. No. 1.35e-10;  
 Matches 112; Conservative 182; Mismatches 303; Indels 37; Gaps 34;  
 DB 888 VNTKPGNNNDLTKRVARDAIANLANLSLETPNVGE-VGIRGGSSQSDSDSAVST 946  
 OY 599 ISNSPGAGIENTRIT-RDGIIFANNWGLDANKPRILPPIAGAGKELTNVOS-AINP 656  
 DB 947 SINMGKID-DSNNKILH-DGTHQSTQEGISLT-AN-THNSEA-TLDRHQTFFHETKG 1000  
 OY 657 ATNGQDLEMRRLSTANKEKSGAATIDLYNSOVPLTFPGAGDPVNTKLSILKVG 716  
 DB 1001 GGOJGVLS-TKTSQSDITVAIKEGQTT--DNALMETKAGSQ-FTSNDGISNNGEN-AHY 1055  
 OY 717 GKTTADDLTKNNIGVAVDSTNSLVLAKTLSDDAVNTKTLTASDKVTVDSNNNAKL 776

DB 1056 EGAQFDAQKCKT-VINAGGDLTLQAQTD--THSRSS-YNVGSANLKVGTTPESKDYGGGF 1112  
 OY 777 QNDLLEFSKONTGAPATNSKITYGVGLKFTDNNNGIALDGTTIYIKKVGAKQ-DGSL 835  
 DB 1113 NAGTHHSKQETAKYGTI-TGSQGIELNAGHNLLOGTHLSEODIALNATNKYDLOSA 1171  
 OY 836 DKSFPYDKRKL-KYGEVETITNGIN-AGKALITGLSTMTLDATNATGHVTLQGLI-VD 891  
 DB 1172 SSEHTEKNNLSGGVOAGEGKKMTDASS-VNGLSAGPALGKODKSVRSREGGITNNG 1230  
 OY 892 SDTKT-BAASIGCVLNAEGLNKNNGDAKDFVSTYDVFDEINGNATKATYVYDGKASVAV 950  
 DB 1231 NLTINGSVHLCQAQVNSKPTOLSSOS-GDIETTSQSDTYKNNMGTDIGF-NGKKTNTT 1288  
 OY 951 DVNDGTTIHLTGADGNKNOIGVKTTLTKTDAGDKALNFVNSGDDKALIAKALIAON 1010  
 DB 1289 PREYEE-KPATSIHNG-ELKLVN-VEDOOKTSHONATLETGT-LTINSKDLTSGAN 1344  
 OY 1011 LNTLAGELRNTKGTADTALQTFQYKKYKENGDDNDADITVVGKDAKTQVNTLKIKGN 1070  
 DB 1345 -VTADSVT-GNVGGSNLIASOKESDRHVTYGVNVGYNHTNDPKSSQVNTAKAGSLLER 1402  
 OY 1071 GLDIQTNKDGTVTFEGINTOSGLKAGNNTLINN-GLSIKNTAGNEQIQ-VG-ADGVKFAK 1127  
 DB 1403 TIKDTISGKSSSTDALSDKNSLSSTIADKTGSDTKAKIIOGFGKVGNGIKNIYVGA 1462  
 OY 1128 VNNGVAGAGIDGTRITRDEIGFAGINGSIDKSPHLSKGINAGKRIIN-IQSEELAQ 1186  
 DB 1463 EGHANADIKVTHVNDVATYKTSLSNNDLSLN 1496  
 OY 1187 NSNDVATGKTYDLKTELENKISSTAKTAQNSLH 1220  
 RESULT 4  
 ID YADA\_YEREN STANDARD; PRT; 455 AA.  
 AC P31489;  
 DT 01-JUL-1993 (REL. 26, CREATED)  
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE INVASIN PRECURSOR (OUTER MEMBRANE ADHESIN).  
 GN YADA OR YOPA OR INVA OR YOP1.  
 OS YERSINIA ENTEROCOLITICA.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 CC ENTEROBACTERIACEAE.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-6471/76 / SEROTYPE O:3;  
 RX MEDLINE; 95020586.  
 RA TAWM A., TARKKANEN A., KORHONEN T.K., KUUSELA P., TOIVANEN P.,  
 RA SKURNIK M.;  
 RU MOL. MICROBIOL. 10:995-1011(1993).  
 CC -1- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO  
 PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE  
 CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS  
 PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL  
 SURFACE.  
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
 DR EMBL; X13882; G48607; -.  
 DR PIR; S04912; S04912.  
 KM PLASMID; VIRULENCE; SIGNAL; OUTER MEMBRANE.  
 FT SIGNAL 1 25  
 FT CHAIN 26 455 INVASIN.  
 SQ SEQUENCE 455 AA; 47136 MW; D1F81809 CRC32;  
 Query Match 1.6%; Score 209; DB 1; Length 455;  
 Best Local Similarity 34.9%; Pred. No. 3.35e-09;  
 Matches 44; Conservative 33; Mismatches 45; Indels 4; Gaps 4;  
 DB 55 PPYPGAGLMAKAKGHSAIGATAEAKGAAYAGASITGVNSVAGISLKGALGDSA 114  
 OY 1825 PVOGGRNGIDSSASGKHSVAIGFAKADGEAAVAVIGROT-QAGNOSIAGDMAQATGQS 1883

Db 115 VTYGAAT-ADKGVATGARAST-SDTGVAVGENSK-ADAKNSVAIGHSSVAAHNGYSI 171  
 Oy 1884 IAIIGNVNAGKHSAGAIIDPSTVAKADNSYSVGNNOFTDATQTDVFGVNNITVTENSV 1943  
 Db 172 AIGDRS 177  
 Oy 1944 ALGNS 1949

RESULT 5  
 ID SLAP\_CACOR STANDARD; PRT; 1025 AA.  
 AC P35828;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
 DE S-LAYER PROTEIN (PARACRYSTALLINE SURFACE LAYER PROTEIN).  
 GN R5A.  
 OS CAULOBACTER CRESCENTUS.  
 NC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; BUDDING AND/OR APPENDAGED.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.  
 RX STRAIN-ATCC 19089 / CB15;  
 RX MEDLINE; 93007489.  
 CC GILCHRIST A., FISHER J.A., SMIT J.;  
 CAN. J. MICROBIOL. 38:193-202(1992).  
 CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY  
 OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIA IS COVERED BY A  
 S-LAYER WITH HEXAGONAL SYMMETRY.  
 CC -1- THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE SURFACE  
 ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A SPECIFIC  
 CC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.  
 DR EMBL; M84760; G289577; -  
 DR HSSP; P22629; 1SRE; CALCIUM-BINDING.  
 KW CELL WALL; S-LAYER; CALCIUM-BINDING.  
 FT INITI MET 0  
 SQ SEQUENCE 1025 AA; 98209 MW; 18938684 CRC32;

Query Match 1.6%; Score 209; DB 1; Length 1025;  
 Best local Similarity 22.3%; Pred. No. 3.35e-09;  
 Matches 191; Conservative 205; Mismatches 389; Indels 70; Gaps 64;

Db 110 NATGAGATAPAAATYGVSYAQTATAYDKIIGNA-VATAGVVAANAVALFSLQANI 168  
 Oy 226 SLAVGLTAFAKA-ASSIAGVSNQAIGFATAVGSGTQVNLKGLAGSGOYLQKNDV 284  
 Db 169 DYLTAVRANTPTTAADIDLAVKAALITITLNAATVSGIGYATATAAMINDLSDGALS 228  
 Oy 285 N-AAVAVRAYAP-DDNQPIDNRKATFKNGADVDSIGNSGNDISIRRIIN-VGAGSAD 340  
 Db 229 TDNAAGVNFETAVPSSVSGSTLSLTGTDLTGANNOTFVAGVAGAAITTVGDTISG 288  
 Oy 341 TD-AVAVVALKKEVRLANQITFK-GDSSNNRVEKGLKTLITTGADQTSALT-DHNGV 397  
 Db 289 GAGTDLVNVQ-AAAVTALPTGVITISGIEYNNVSGAA-ITLNTSSGYGLTALNTS- 345  
 Oy 398 YONGGLK-VOLAERLISLKVTEENLPAKEVYGTGRRLTDKIGFINDMNGIDESKPY 456  
 Db 346 GAAQVTTAGGNLTATTAQAANVAVDGRANVYASTGVTSCTTYGANSASAGTVSV 405  
 Oy 457 LKDDGIHNGGKIKRLTAGVDDDAAYGQLK-KVNOTA-ESALQPTTVKVKNG-N- 512  
 Db 406 SVANSSTTTGAIAVGTAV-TVAQTAGNANVTLLTQADVT-TGNSSTAVVYVQTAA 463  
 Oy 513 DANDSKITVGNKNNPDGQVNTLKLKGENGVDT-TETNGTVTFGLQNNGLTVGNSTL 571  
 Db 464 ATAGTAVAGRVNGAVTTIDSAASATAGKATAYVLSFGAATIDSSALTYVNLSTGTS 523  
 Oy 572 NNDGSLVKN-TSNKQI-OVGADGITFD-ISNSKPA-GIENTRIRRDIGFANNNGS 627  
 Db 524 LGIGGATLAPPT-ANTLTL-TVNLITTTGALTDEAADAAGFTITINAGSTASTASL 581  
 Oy 628 LDANKRLPTGIGINAGKELTNVOS-AINPATNGQLDPMNRLSTANTERKSGSAATIKDL 686

Db 582 VAADATTLINSGDANVTITSHAAALITGITVNSVATLGAELATGLVFTGAGRDSTLL 641  
 Oy 687 YNLSOVLTFEADTGPNTVKLIGELK-VKGGKTTADDLTKN-NIGVV-ADST-NS-LT 741  
 Db 642 GATTAIVAGAGDDTVTVSSATLGAAGSVNGGDDTDVAVANVNGSSFPADPFGFTLR 701  
 Oy 742 VKLAKT--LSDLD-AVNKRITLADSKYTVDSGNTAKL-QNGD-LITSKONT-GA--TPA 793  
 Db 702 VAGAAAGSHNANGFTALQGTAGATTFTVAVNVGILTVLAPGTITVLANATGSTD 761  
 Oy 794 TNSKITVYDGLK-FTDNN-GIALDGTYYIKD-KVGFAGKDGSLDKSPYLDKDK-LK- 848  
 Db 762 VENTLSSAALAGTVALAGETVNAATDNTAHVDITL-QANS-AKSIYVTS-N 817  
 Oy 849 VGEVEITING-INAGKAITGSLNLTDAINA-TTGHVYQGLVDSYDKRPAASIGDVLN 906  
 Db 818 AGNLINTNGNT-A-VTSFASA-VTGAP-A-VTFVSANTVGEVYIRGAGADSLTGS 872  
 Oy 907 AGFNLKNGDAKDFVSTVDYDFINGNATFAKYTDGKASKVAVDANV-DGTTIH-LTG- 963  
 Db 873 ATANDTIIGAGADTVYTGTDFTGTGADIPDIAIGTSTAFTTDAVAGDKID-L 931  
 Oy 964 ADGNKNQIG-VKTTL--TK-TDA-KQDKALIN-FSVNS-GDDKALINAKD--IADNINTL 1014  
 Db 932 VG-ISTNGAIDGAF 945  
 Oy 1015 AGEIRNTKGTADTL 1029

RESULT 6  
 ID YADA YERPS STANDARD; PRT; 434 AA.  
 AC P10858;  
 DT 01-JUL-1989 (REL. 11, CREATED)  
 DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)  
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)  
 DE INVASIN PRECURSOR (OUTER MEMBRANE ADHESIN).  
 GN YADA OR YOPA OR INVA OR YOP1.  
 OS YERSINIA PSEUDOTUBERCULOSIS.  
 CC PLASMID PIB1.  
 CC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 CC ENTEROBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-YPIII;  
 RX MEDLINE; 88302441.  
 RA ROSOVIST R., SKURNIK M., WOLF-WATZ H.;  
 RL NATURE 334:522-525(1988).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN-YPIII;  
 RX MEDLINE; 89343638.  
 RA SKURNIK M., WOLF-WATZ H.;  
 RL MOL. MICROBIOL. 3:517-529(1989).  
 CC -1- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO  
 PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE  
 CC CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS  
 CC PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL  
 CC SURFACE.  
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
 DR EMBL; X13883; G48640; -  
 DR PIR; S04534; S04534.  
 DR PIR; S04910; S04910.  
 KW PLASMID; VIRULENCE; SIGNAL; OUTER MEMBRANE.  
 FT SIGNAL 1  
 FT CHAIN 26  
 SQ SEQUENCE 434 AA; 45054 MW; 73DC3575 CRC32;

Query Match 1.5%; Score 207; DB 1; Length 434;  
 Best local Similarity 24.7%; Pred. No. 5.68e-09;  
 Matches 65; Conservative 80; Mismatches 107; Indels 11; Gaps 10;

Db 142 SKALGSANTYAGASSTAGKDG-VALGARASASDITGVAVGENSKYDAQNSVAIGHSHVAA 200



01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE APOMICIN (MUCIN CORE PROTEIN) (FRAGMENT).  
 OS SUS SCROFA (PIG).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN EUTHERIA; ARTIODACTYLA.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-SUBMAXILLARY GLAND;  
 RX MEDLINE; 91236743.  
 RA ECKHARDT A.E., TIMPTE C.S., ABERNETHY J.L., ZHAO Y., HILL R.L.;  
 RL J. BIOL. CHEM. 266:9678-9686(1991).  
 [2]  
 RP SEQUENCE OF 1-503 FROM N.A.  
 RC TISSUE-SUBMAXILLARY GLAND;  
 RX MEDLINE; 88087170.  
 RA TIMPTE C.S., ECKHARDT A.E., ABERNETHY J.L., HILL R.L.;  
 RL J. BIOL. CHEM. 263:1081-1088(1988).  
 [3]  
 RP SEQUENCE OF 45-80.  
 RC TISSUE-SUBMAXILLARY GLAND;  
 RX MEDLINE; 87280230.  
 RA ECKHARDT A.E., TIMPTE C.S., ABERNETHY J.L., TOWNADJE A.,  
 RA JOHNSON W.C. JR., HILL R.L.;  
 RL J. BIOL. CHEM. 262:11339-11344(1987).  
 [4]  
 CC CARBOHYDRATE-BINDING SITES.  
 RA GERKEN T.A., OWENS C.L., PASUMARTHY M.;  
 RL J. BIOL. CHEM. 272:9709-9719(1997).  
 CC -1- FUNCTION: APOMICIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN  
 SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY  
 GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SEVE TO LUBRICATE  
 THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL  
 ENVIRONMENT.  
 CC -1- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A  
 MULTIMERIC MUCIN STRUCTURE.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.  
 CC -1- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81  
 RESIDUES.  
 CC -1- PTM: CONTAINS MANY POTENTIAL SITES FOR O-LINKED GLYCOSYLATION.  
 CC -1- SIMILARITY: CONTAINS A VWFC DOMAIN.  
 CC -1- SIMILARITY: CONTAINS A C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).  
 DR EMBL; M61883; G1634374; .  
 DR EMBL; M21174; E11628; ALT\_SEQ.  
 DR PIR; A40009; A40009.  
 DR HSSP; P22629; 1SRF.  
 DR PROSITE; PS01185; CTCK\_1; 1.  
 DR PROSITE; PS01208; VWFC; 1.  
 DR PROSITE; PS01225; CTCK\_2; 1.  
 DR GLYCOPROTEIN; REPEAT.  
 ET NON\_TER 1  
 REPEAT <1 67  
 REPEAT 68 148  
 REPEAT 149 229  
 REPEAT 230 310  
 REPEAT 311 391  
 REPEAT 929 995  
 DOMAIN 1062 1146  
 DOMAIN 1062 1109  
 DISULFID 1076 1123  
 DISULFID 1085 1139  
 DISULFID 1089 1141  
 DISULFID 1145 1145  
 ? 1145 46  
 CARBOHYD 46 50  
 CARBOHYD 50 51  
 CARBOHYD 51 51  
 CARBOHYD 57 57  
 CARBOHYD 58 58  
 CARBOHYD 61 61  
 CARBOHYD 66 66  
 CARBOHYD 67 67  
 CARBOHYD 73 73

ET CARBOHYD 74 74  
 ET CARBOHYD 76 76  
 ET CARBOHYD 77 77  
 ET CARBOHYD 81 81  
 ET CARBOHYD 83 83  
 ET CARBOHYD 87 87  
 ET CARBOHYD 91 91  
 ET CARBOHYD 93 93  
 ET CARBOHYD 94 94  
 ET CARBOHYD 96 96  
 ET CARBOHYD 98 98  
 ET CARBOHYD 101 101  
 ET CARBOHYD 103 103  
 ET CARBOHYD 104 104  
 ET CARBOHYD 106 106  
 ET CARBOHYD 107 107  
 ET CARBOHYD 108 108  
 ET CARBOHYD 110 110  
 ET CARBOHYD 114 114  
 ET CARBOHYD 117 117  
 ET CARBOHYD 123 123  
 ET CARBOHYD 124 124  
 ET CARBOHYD 418 418  
 ET CARBOHYD 547 547  
 ET CARBOHYD 917 917  
 ET CARBOHYD 985 985  
 ET CARBOHYD 1002 1002  
 ET CARBOHYD 1068 1068  
 SQ SEQUENCE 1150 AA; 109615 MM; F7C55GCE CRC32;  
 Query Match 1.4%; Score 186; DB 1; Length 1150;  
 Best Local Similarity 22.4%; Pred. No. 1.32e-06;  
 Matches 67; Conservative 90; Mismatches 128; Indels 14; Gaps 12;  
 Db 209 VAGSSGAPAVSSGA-SOAGTSGAGPCTTASVGV-TETARPVAGSGTGTGVSAGSGT 266  
 QY 1827 VQGRNGIDSSAGKSHVAIGFQAKADEAAVAIGROTQAGNOSIAIGDNQATGDOSIAI 1886  
 Db 267 GSSSGSPGATGASIGQETSRISVAGSSGAPAVSSGAQAGTSGAGPCTTAS-SVGYTE 325  
 QY 1887 GTGNVAGKHSIGAIGDPSTYKADNSYVGNNGFTDALTQD-VFVGNNITVIESNVSL 1945  
 Db 326 TARSVAGSGTGTGVSAGSGTSSSSSPATGATGASIGQETSRISVAGSSGAPAVSSGAS 385  
 QY 1946 GSNASISAGTHAGTQAKKSDGTTTATGATGTVGFAGQITAVG-AVSYGASGAEARRIO 2004  
 Db 386 QAGTSEATTSIEGAGTSGVGFK-TEATTPGEMETTRV-GIATGTGIVSKTLE--P 440  
 QY 2005 NVAAGEVSATSTDAVNSQL-YKATQCIANATNELDHRIONENKANAGISSAMAMASMP 2063  
 Db 441 GSYTEATTSIGRSCTTDLPGGTIVLPFSSHSSQS-KPGSSVYTPGSPSGSETG 498  
 QY 2064 QAYLPGRSMTVG-GIATVHNG-QGAVAVAGLSKLSDNQWFKINGSADTQGHVCAAVGAG 2120  
 RESULT 10  
 ID FLAG\_SERMA STANDARD; PRT; 351 AA.  
 AC P13713;  
 DT 01-JAN-1990 (REL. 13, CREATED)  
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE FLAGELLIN.  
 GN FLIC OR FLAP OR HNG.  
 OS SERRATIA MARCESCENS.  
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
 RN ENTEROBACTERIACEAE.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-274.  
 RX MEDLINE; 89378766.  
 RA HARSHNEY R.M., ESTEPA G., YANAGI H.;  
 RL GENE 79:1-8(1989).  
 CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO

1

Downloaded from <http://ajphaphysocpharm.sagepub.com/> at 11:04 11 November 2014



ID	PER_DROME	STANDARD;	PRT;	1155 AA.
Db	141	SSASADSVSSVSSSSSSASDVSSVSQSAKSAKSDVSSVSQDSASSTSDVSSVSQSSSSA	200	
Qy	1890	NVYAGKHSIGADGPSTYKADNSISYGVNNNOFTDATQTDVFGVGNNTI-VLESNSVALGSN	1948	
Db	201	SDVSSVSQAS-SASDVSSVSQASASTSDVSSVSQASASTSGVSSSGSQS-VSS-AS	257	
Qy	1949	SATAGTHAGTQAKKSDSGTGTTTGTGATGYKGFAGGTAVGAVSGAERRIQWVYA	2008	
Db	258	GS-SSSPQSTSSASTASGS-ATNSLSLSTSSASSASASTASNSLSSDDGTYLPPTTIS	315	
Qy	2009	GEVSATSTDAVNGSQLTKATQGIANTNTELDRIHQENKNAAGSISAMAMSPQAYIP	2068	
Db	316	GDLTLTGKVIATG-VVYAG-AKLT	339	
Qy	2069	GRSMVTGGIATHNGGAVAVGLSKLS	2094	
Db	15	PER_DROME	STANDARD;	PRT;
AC	P07663;			
Dt	01-APR-1988	(REL. 07, CREATED)		
Dt	01-APR-1988	(REL. 07, LAST SEQUENCE UPDATE)		
Dt	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)		
Dt		PERIOD CLOCK PROTEIN.		
OS	PER.			
OC	DROSOPHILA MELANOGASTER (FRUIT FLY).			
OC	EURARVOTA, METAZOA; ARTHROPODA; INSECTA; DIPTERA.			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CANTON-S.			
RA	MEDLINE; 86146900.			
RL	JACKSON F.R., BARGIELLO T.A., YUN S.-H., YOUNG M.W.;			
RL	NATURE 320:185-188(1986).			
RP	[2]			
RP	SEQUENCE OF 436-1006 FROM N.A.			
RC	STRAIN-OREGON-R.			
RA	MEDLINE; 86245055.			
RL	REDDY P., JACQUIER A.C., ABOVICH N., PETERSEN G., ROSASH M.;			
RL	CELL 46:53-61(1986).			
RP	[3]			
RP	SEQUENCE OF 1-511 FROM N.A.			
RX	MEDLINE; 93170641.			
RA	KLIEMAN R.W., HEY J.;			
RA	GENETICS 133:375-387(1993).			
CC	-I- FUNCTION: INVOLVED IN THE GENERATION OF BIOLOGICAL RHYTHMS.			
CC	EXHIBITS A CIRCADIAN OSCILLATION IN ABUNDANCE. INTERACTS WITH THE			
CC	TIMELESS PROTEIN (TIM).			
CC	-I- SUBCELLULAR LOCATION: NUCLEAR. FIRST ACCUMULATES IN THE			
CC	PERINUCLEAR REGION ABOUT AN HOUR BEFORE IT IS TRANSPORTED TO THE			
CC	NUCLEUS.			
CC	-I- DOMAIN: THE SEQUENCE CONTAINS A REMARKABLE RUN OF ALTERNATING GLY			
CC	THR RESIDUES. THE LENGTH OF THE G-T REGION IS SIMILAR TO THAT OF			
CC	THE MAMMALIAN G-S REPEAT IN A CHONDRITIN SULFATE PROTEOGLYCAN.			
CC	BIOCHEMICAL CHARACTERIZATION OF THIS ANTIGEN INDICATES THAT IT IS			
CC	A PROTEOGLYCAN.			
CC	-I- SIMILARITY: TO THE SINGLE-MINDED (SIM), ARNT AND AH RECEPTOR			
CC	PROTEINS.			
DR	EMBL; X03636; G8329; -.			
DR	EMBL; M11969; G158055; -.			
DR	EMBL; D00009; -, NOT_ANNOTATED_CDS.			
DR	EMBL; L07817; G552118; -.			
DR	EMBL; L07819; G552117; -.			
DR	EMBL; L07819; G552116; -.			
DR	EMBL; L07821; G552114; -.			
DR	EMBL; L07823; G552112; -.			
DR	EMBL; L07825; G552110; -.			
DR	EMBL; L07825; G552110; -.			
DR	PIR; A23932; UMF.			
DR	FLYBASE; FBgn0003068; per.			
KW	BIOLOGICAL RHYTHMS; GLYCOPROTEIN; PROTEOGLYCAN; REPEAT;			
KW	NUCLEAR PROTEIN.			
FT	REPEAT	186		
FT	REPEAT	336		
FT	REPEAT	387		
FT	DOMAIN	632		
FT		236		
FT		387		
FT		675		
FT		PAS IRL.		
FT		PAS IRL.		
FT		G-T REPEATS.		

[illegible]

Search completed: Wed Oct 28 11:12:57 1998  
Job time : 244 secs.

